

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ  
імені ІГОРЯ СІКОРСЬКОГО»**

**Інститут прикладного системного аналізу  
Кафедра математичних методів системного аналізу**

«На правах рукопису»  
УДК 004.415.2; 519.246.8

«До захисту допущено»

Завідувач кафедри

\_\_\_\_\_ О. Л. Тимошук

«\_\_» \_\_\_\_\_ 20\_\_ р.

**Магістерська дисертація**

**на здобуття ступеня магістра**

**зі спеціальності 124 Системний аналіз**

**на тему: «Моделі для прогнозування характеристик трафіка цифрової  
реклами»**

Виконав:

студент II курсу, групи КА-61м

Логін Вадим Вікторович \_\_\_\_\_

Науковий керівник:

професор кафедри ММСА ІПСА

КПІ ім. Ігоря Сікорського

д.т.н., проф., Бідюк П. І. \_\_\_\_\_

Рецензент:

професор кафедри інформаційної безпеки

фізико-технічного інституту

КПІ ім. Ігоря Сікорського,

д.т.н., проф., Архипов О. Є. \_\_\_\_\_

Засвідчую, що у цій магістерській  
дисертації немає запозичень з праць  
інших авторів без відповідних посилань.  
Студент \_\_\_\_\_

Київ  
2018

**Національний технічний університет України  
«Київський політехнічний інститут імені Ігоря Сікорського»**

**Інститут прикладного системного аналізу**

**Кафедра математичних методів системного аналізу**

Рівень вищої освіти – другий (магістерський)

Спеціальність (спеціалізація) – 124 «Системний аналіз» («Системний аналіз і управління»)

ЗАТВЕРДЖУЮ

Завідувач кафедри

\_\_\_\_\_ О. Л. Тимошук

«\_\_» \_\_\_\_\_ 20\_\_ р.

**ЗАВДАННЯ  
на магістерську дисертацію студенту  
Логіну Вадиму Вікторовичу**

1. Тема дисертації «Моделі для прогнозування характеристик трафіка цифрової реклами», науковий керівник дисертації Бідюк Петро Іванович, д.т.н., проф., затверджені наказом по університету від «\_\_» \_\_\_\_\_ 20\_\_ р. № \_\_\_\_\_
2. Термін подання студентом дисертації \_\_\_\_\_
3. Об'єкт дослідження: трафік цифрової реклами у формі статистичних даних.
4. Предмет дослідження: моделі та методи аналізу даних у формі часових рядів, методи прикладної статистики.
5. Перелік завдань, які потрібно розробити: (1) виконати загальний огляд моделей часових рядів та вибрати ті, які відповідають поставленій задачі прогнозування характеристик трафіка цифрової реклами; (2) побудувати моделі часових рядів за допомогою обраних підходів; (3) порівняти отримані моделі та вибрати найкращу; (4) виробити рекомендації щодо вдосконалення у майбутньому розроблених підходів до розв'язання даної задачі.
6. Орієнтовний перелік графічного (ілюстративного) матеріалу: (1) структура ринку цифрової реклами; (2) структура вихідних статистичних даних; (3) загальна блок-схема процесу моделювання; (4) результати обчислювальних експериментів.

7. Орієнтовний перелік публікацій: стаття «Models for forecasting parameters of digital advertising traffic» у видавництві CEUR, тези «Моделі для прогнозування характеристик трафіка цифрової реклами» на конференції САІТ-2018.

8. Дата видачі завдання \_\_\_\_\_

#### Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації	Примітка
1	Оформлення концептуального вступу	12.03.2018 – 18.03.2018	
2	Оформлення першого розділу магістерської дисертації	19.03.2018 – 25.03.2018	
3	Оформлення другого розділу магістерської дисертації	26.03.2018 – 01.04.2018	
4	Робота над статтею за результатами наукового дослідження	02.04.2018 – 08.04.2018	
5	Оформлення третього розділу магістерської дисертації	09.04.2018 – 22.04.2018	
6	Завершення роботи над основною частиною магістерської дисертації; оформлення розділу щодо розроблення стартап-проекту	23.04.2018 – 06.05.2018	
7	Оформлення висновків і графічної частини магістерської дисертації	07.05.2018 – 13.05.2018	
8	Представлення роботи до захисту	14.05.2018 – 16.05.2018	

Студент

В. В. Логін

Науковий керівник дисертації

П. І. Бідюк

## РЕФЕРАТ

Магістерська дисертація: 112 с., 48 рис., 40 табл., 3 додатки і 30 джерел.

Об'єкт дослідження – трафік цифрової реклами у формі статистичних даних.

Предмет дослідження – моделі та методи аналізу даних у формі часових рядів, методи прикладної статистики.

Мета роботи – побудова моделей часових рядів для прогнозування найважливіших характеристик трафіка цифрової реклами.

Методи дослідження – моделі часових рядів для прогнозування даних та порівняльний аналіз отриманих моделей.

У даній роботі наведені результати побудови моделей часових рядів, що призначені для прогнозування найважливіших характеристик трафіка цифрової реклами. Описані результати порівняльного аналізу отриманих моделей за допомогою інформаційних критеріїв, а також з точки зору їхньої точності. Встановлено, що для нашої задачі, найкращою моделлю є модель ARIMAX (Autoregressive integrated moving-average model with exogenous inputs), тобто модель авторегресії та ковзного середнього з екзогенними змінними. Тому для подальших досліджень рекомендовано використовувати саме цю модель.

За матеріалами магістерської дисертації були написані тези, а також написана наукова стаття. Тези будуть опубліковані в збірці тез доповідей конференції САІТ-2018. А наукова стаття буде опублікована в електронній збірці доповідей у видавництві CEUR.

Прогнозні припущення щодо подальшого розвитку об'єкта дослідження – побудова нових, а також вдосконалення існуючих моделей часових рядів для прогнозування найважливіших характеристик цифрової реклами. А також узагальнення дослідження, що проводилось у даній роботі, на аналіз окремих сайтів із рекламного трафіку.

ТРАФІК ЦИФРОВОЇ РЕКЛАМИ, СТАТИСТИЧНИЙ АНАЛІЗ ДАНИХ, ЧАСОВІ РЯДИ, ПРОГНОЗУВАННЯ ДАНИХ У ФОРМІ ЧАСОВИХ РЯДІВ, МОДЕЛЬ AR, МОДЕЛЬ ARMA, МОДЕЛЬ ARIMA, МОДЕЛЬ ARIMAX.

## ABSTRACT

Models for forecasting parameters of digital advertising traffic.

Master's thesis: 112 p., 48 fig., 40 tabl., 3 appendixes and 30 sources.

The object of study – digital advertising traffic in the form of statistical data.

Subject of research – models and methods of analysis of data in the form of time series, methods of applied statistics.

Purpose – constructing time series models for forecasting the most important characteristics of digital advertising traffic.

Methods of research – time series models for forecasting data and comparative analysis of the obtained models.

This paper presents the results of construction of time series models, which are intended for forecasting of the most important characteristics of digital advertising traffic. Described the results of the comparative analysis of the obtained models with the help of information criteria, and also in terms of their accuracy. Was found that for our task, the best model is the ARIMAX model (Autoregressive integrated moving-average model with exogenous inputs). Therefore, it is recommended to use this model for further research.

Based on master's dissertation were written theses as well as a scientific article. The theses will be published in the SAIT-2018 conference Book of Abstracts. The scientific article will be published in the electronic collection of reports at the CEUR publishing house (CEUR Workshop Proceedings).

The further development of the research object – is the construction of new ones, as well as the improvement of existing time series models for forecasting the most important characteristics of digital advertising traffic. And also – it is a generalization of the research, conducted in this paper, on the analysis of individual sites from the digital advertising traffic.

DIGITAL ADVERTISING TRAFFIC, STATISTICAL ANALYSIS OF DATA, TIME SERIES, FORECASTING OF DATA IN THE FORM OF TIME SERIES, AR MODEL, ARMA MODEL, ARIMA MODEL, ARIMAX MODEL.

## ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ.....	8
ВСТУП.....	10
РОЗДІЛ 1 ЦИФРОВА РЕКЛАМА ТА ПРОГНОЗУВАННЯ ХАРАКТЕРИСТИК ЇЇ ТРАФІКА.....	13
1.1 Актуальність прогнозування характеристик трафіка цифрової реклами.....	13
1.2 Поняття цифрової реклами та її задачі.....	14
1.3 Трансформація цифрової реклами з часом, її типи та види.....	18
1.3.1 Трансформація цифрової реклами в часі.....	18
1.3.2 Типи цифрової реклами.....	23
1.3.3 Види цифрової реклами.....	27
1.4 Огляд математичних методів для моделювання і прогнозування характеристик трафіка цифрової реклами.....	34
Висновки до розділу 1 та постановка задачі дослідження.....	37
РОЗДІЛ 2 МЕТОДИКА МОДЕЛЮВАННЯ ТА ПРОГНОЗУВАННЯ ХАРАКТЕРИСТИК ТРАФІКА ЦИФРОВОЇ РЕКЛАМИ.....	39
2.1 Аналіз та конкретизація поставлених задач.....	39
2.2 Методика побудови математичних моделей на основі статистичних даних у формі часових рядів.....	40
2.3 Вибір моделей для прогнозування характеристик трафіка цифрової реклами.....	48
2.3.1 Модель AR.....	49
2.3.2 Модель ARMA.....	51
2.3.3 Модель ARIMA.....	55
2.3.4 Модель ARIMAX.....	57
2.4 Критеріальна база для аналізу якості результатів та вибору кращої моделі.....	58
Висновки до розділу 2.....	64
РОЗДІЛ 3 ПОБУДОВА МОДЕЛЕЙ ПРОГНОЗУ ВАЖЛИВИХ ХАРАКТЕРИСТИК ТРАФІКА ЦИФРОВОЇ РЕКЛАМИ.....	65
3.1 Вибір даних для обробки та формулювання конкретних задач.....	65
3.1.1 Опис статистичний даних для прогнозування.....	65

	7
3.1.2 Постановка конкретних задач для проведення моделювання .....	68
3.1.3 Блок-схема розв'язання поставленої задачі.....	73
3.2 Попередня обробка вхідних даних .....	73
3.3 Вибір значущих параметрів.....	76
3.4 Розбиття вихідної вибірки на навчальну та тестову .....	79
3.5 Побудова моделей прогнозу доходу та заповнюваності .....	80
3.5.1 Побудова моделі множинної регресії.....	80
3.5.2 Побудова моделі AR .....	82
3.5.3 Побудова моделі ARMA.....	86
3.5.4 Побудова моделі ARIMA.....	88
3.5.5 Побудова моделі ARIMAX.....	91
3.6 Вибір кращої моделі прогнозу для доходу та заповнюваності.....	93
Висновки до розділу 3 .....	96
РОЗДІЛ 4 РОЗРОБЛЕННЯ СТАРТАП-ПРОЕКТУ .....	98
4.1 Опис ідеї проекту.....	98
4.2 Технологічний аудит ідеї проекту .....	100
4.3 Аналіз ринкових можливостей запуску стартап-проекту .....	100
4.4 Розроблення ринкової стратегії проекту.....	104
4.5 Розроблення маркетингової програми стартап-проекту .....	105
Висновки до розділу 4 .....	107
ВИСНОВКИ ЗА РОБОТОЮ ТА ПЕРСПЕКТИВИ ДЛЯ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ.....	108
ПЕРЕЛІК ПОСИЛАНЬ .....	110

## ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

- SMS – Short Message Service (служба коротких повідомлень)
- MMS – Multimedia Messaging Service (служба мультимедійних повідомлень)
- CTR – Click-Through Rate (рейтинг кліків або клікабельність)
- SAR – Secondary Action Rate (відсоток показів реклами, які трансформувались у певну подальшу дію, після початкового кліку по рекламі)
- SEO – Search Engine Optimization (пошукова оптимізація сайту)
- SEM – Search Engine Marketing (пошуковий маркетинг)
- PPC – Pay-Per-Click (реklamна модель «платити за клік»)
- SNS – Simple Notification Service (служба простих сповіщень)
- PR – Public Relations (зв'язки з громадськістю)
- AR – Autoregressive model (модель авторегресії)
- MA – Moving-average model (модель ковзного середнього)
- ARMA – Autoregressive moving-average model (модель авторегресії та ковзного середнього)
- ARIMA – Autoregressive integrated moving average model (модель авторегресії та інтегрованого ковзного середнього)
- ARIMAX – Autoregressive Integrated Moving Average model with Explanatory Variables (модель авторегресії та інтегрованого ковзного середнього з екзогенними змінними)
- ACF – Auto Correlation Function (автокореляційна функція)
- PACF – Partial Auto Correlation Function (часткова автокореляційна функція)
- CMO – Chief Marketing Officer (директор із маркетингу)
- Ad Ops – Advertising Operations (операції щодо підтримки цифрової реклами)
- SSP – Supply-Side Platform (програмна платформа для автоматизованого продажу рекламного трафіка)
- DSP – Demand-Side Platform (програмна платформа для автоматизованої купівлі рекламного трафіка)
- DMP – Data Management Platform (платформа для управління даними)



$R^2$  – коефіцієнт множинної детермінації

DW – Durbin-Watson statistic (статистика Дарбіна-Уотсона)

AIC – Akaike Information Criterion (інформаційний критерій Акайке)

BIC – Bayesian Information Criterion (інформаційний критерій Байєса)

RMSE – Root Mean Square Error (середньоквадратична похибка)

MAE – Mean Absolute Error (середня абсолютна похибка)

MAPE – Mean Absolute Percentage Error (середня абсолютна похибка у відсотках)

Q – статистика для масштабування

MASE – Mean Absolute Scaled Error (середня абсолютна масштабована похибка)

MBE – Mean Bias Error (середня зміщена похибка)

SMAPE – Symmetric Mean Absolute Percentage Error (симетрична середня абсолютна похибка у відсотках)

## ВСТУП

Однією з поширених сфер, у якій потрібна побудова моделей прогнозу певних характеристик, а особливо тих, що мають безпосереднє відношення до прибутку та успішності самого бізнесу є область цифрового маркетингу (з англ. digital marketing) – маркетинг продуктів або послуг з використанням цифрових каналів для охоплення споживачів [1]. Головна мета – просування брендів через різні форми цифрових медіа. Цифровий маркетинг поширюється за межі інтернет-маркетингу, щоб включати канали, які не вимагають використання Інтернету. Вона включає в себе мобільні телефони (як SMS, так і MMS), маркетинг в соціальних мережах, медійна реклама, маркетинг в пошукових системах та будь-яка інша форма цифрових засобів масової інформації. Більшість експертів вважають, що "цифровий маркетинг" – це не просто ще один канал для маркетингу. Це вимагає нового підходу до маркетингу та нового розуміння поведінки клієнтів. Наприклад, це вимагає, щоб компанії аналізували та кількісно оцінювали завантаження додатків на мобільних пристроях, твіти у Twitter, вподобання на Facebook і так далі [2].

Одним із прикладів успішної кампанії в галузі цифрових медіа була рекламна кампанія створена компанією Pizza Hut. Вони створили програму, яка дозволила клієнтам створювати власну піцу, перетягуючи обрані елементи начинки на графічне зображення основи для піци. Також ця програма могла визначати, який з тисяч закладів даної мережі є найближчим до клієнта. Компанія рекламувала новий додаток в Інтернеті, в друкованому вигляді та на телебаченні – навіть виграла місце у власному комерційному магазині Apple. Протягом двох тижнів програма від Pizza Hut була завантажена 100 000 разів, і протягом трьох місяців користувачі iPhone замовили піцу на суму 1 млн. дол. США. Додаток тепер має мільйони користувачів на платформах iPhone, iPad та Android.

Тому можна із впевненістю сказати, що задачі цифрового маркетингу в цілому, і зокрема задачі прогнозування найважливіших характеристик трафіка

цифрової реклами – є дуже актуальними, оскільки цифрова реклама є досить поширеною в наш час, причому її правильна організація за наявності математичного обґрунтованого прогнозування та планування, може дозволити бізнесу залучити велику кількість нових клієнтів, утримати наявних, а також збільшити розмір свого прибутку.

Дослідженням сутності та розробкою методик цифрової реклами займалися такі спеціалісти у цій області як Д. Майкл, Л. Фредрік, С. Террі [3], а також – Д. Д. Шенбахлер, Д. Л. Гордон, Д. Фолі, Л. Спеллман [4]. Вагомий внесок у дослідження питань ефективності та прибутковості цифрової реклами зробили наступні спеціалісти з маркетингу: К. Харт, Н. Доєрті, Ф. Елліс-Чадвік, М. Кейтс [5, 6] та інші.

Аналіз тематичної літератури різних авторів, котра присвячена поняттю «цифрового маркетингу», свідчить про те що у даному напрямку напрацьована досить велика кількість практичних результатів. Однак варто зазначити, що у різних методичних підходах до цифрової реклами, є свої переваги та недоліки. Це значною мірою ускладнює процес вибору певного підходу, хоча більшість основних підходів до проведення цифрового маркетингу зазвичай забезпечують позитивний результат.

Оскільки ХХІ століття – це ера інформаційного суспільства, тому останнім часом існує необхідність застосування інформаційних технологій та нових підходів до всіх напрямків діяльності, зокрема до маркетингу. Одним із пріоритетних засобів при проведенні сучасного цифрового маркетингу є моделювання процесів, що у ньому відбуваються, зокрема із використанням часових рядів.

Дослідження, що описується у даній роботі, присвячене прогнозуванню найважливіших характеристик трафіку цифрової реклами. Зокрема наведено опис побудови моделей часових рядів для прогнозування конкретних параметрів, а також процедура вибору кращої моделі із побудованих.

Пояснювальна записка до магістерської дисертації складається з чотирьох розділів. У першому розділі описується поняття цифрової реклами та потреба у

прогнозуванні характеристик її трафіка. Другий розділ присвячений вибору моделей для прогнозування параметрів цифрового рекламного трафіку. У третьому розділі наведено опис процесу побудови моделей прогнозу. Також здійснюється порівняння побудованих моделей та даються рекомендації щодо вибору кращої моделі. У четвертому розділі розглянуто питання розроблення стартап-проекту за тематикою магістерської дисертації.

## РОЗДІЛ 1 ЦИФРОВА РЕКЛАМА ТА ПРОГНОЗУВАННЯ ХАРАКТЕРИСТИК ЇЇ ТРАФІКА

### 1.1 Актуальність прогнозування характеристик трафіка цифрової реклами

Як вже зазначалось у вступі, досить поширеною сферою, у якій потрібне прогнозування характеристик та порівняння моделей, є цифровий маркетинг – широкий термін, що стосується багатьох та різних рекламних технологій, що використовуються для охоплення клієнтів за допомогою цифрових технологій. Цифровий маркетинг втілюється широким вибором тактик з обслуговування, продукту та маркетингу, в основному використовуючи Інтернет як основне рекламне середовище, а додатковими середовищами є мобільні телефони та традиційні телебачення і радіо. Цифровий маркетинг також відомий як Інтернет-маркетинг, але їх фактичні процеси відрізняються, оскільки цифровий маркетинг вважається більш цілеспрямованим, вимірюваним та інтерактивним [7]. Цифровий маркетинг включає в себе безліч методів інтернет-маркетингу, таких як оптимізація для пошукових систем (SEO), пошуковий маркетинг (SEM) та створення посилань. Він також поширюється на канали, які не вимагають доступу до Інтернет, що забезпечують цифрові середовища, такі як сервіс коротких повідомлень (SMS), послуги мультимедійних повідомлень (MMS), зворотні виклики та мобільні мелодії на дзвінок, електронні книги, оптичні диски та ігри.

Ключовою задачею цифрового маркетингу є залучення клієнтів та надання їм можливості взаємодіяти з брендом за допомогою обслуговування та постачання цифрових носіїв. Це досягається шляхом проектування цифрових засобів масової інформації таким чином, що для цього потрібен певний тип дій кінцевого користувача для перегляду або отримання основного задуму, задля якого дана реклама була створена. Наприклад, щоб отримати безкоштовну електронну книгу, клієнту може бути потрібно зареєструвати або заповнити форму, забезпечуючи рекламодавця цінним клієнтом або «лідом», тобто

потенційним клієнтом, який входить до цільової аудиторії певної компанії та при цьому він зацікавлений у конкретному товарі або послугі.

У цій ситуації має бути проведений аналіз історичних даних рекламного трафіка. Потрібно проаналізувати та спрогнозувати ключові характеристики трафіка цифрової реклами для подальшого прийняття відповідних рішень. Задача аналітика полягає у тому, що він має виділити набір таких ключових характеристик цифрового рекламного трафіка та побудувати прогностичні моделі, які будуть визначати динаміку змін обраних головних характеристик трафіка. Знаючи, наприклад, прогнози параметрів доходу та заповнюваності рекламного трафіка, можна планувати власні витрати, розмірковувати над розширенням трафіка, його покращенням або оптимізацією. Також на основі цих даних можна робити висновки щодо певних довгострокового партнерства або диверсифікації трафіка цифрової реклами.

Отже, можна зробити висновок, що задача прогнозування ключових характеристик трафіка цифрової реклами є досить актуальною у наш час, оскільки цифрова реклама є достатньо поширеною, до того ж отримані прогнози дають додаткову, математично обґрунтовану, інформацію, яку можна врахувати при здійсненні стратегічного планування.

## 1.2 Поняття цифрової реклами та її задачі

Цифрова реклама – це використання Інтернету, мобільних пристроїв, соціальних мереж, пошукових систем, медійної реклами та інших каналів для охоплення споживачів. Як підмножина традиційного реклами, цифрова реклама виходить за рамки інтернету, та включає в себе служби коротких повідомлень (SMS), просту службу сповіщення (SNS), оптимізацію для пошукових систем (SEO), електронні чи інтерактивні рекламні щити та іншу онлайн-рекламу (наприклад, банерні об'яви) для рекламування товарів чи послуг. Деякі фахівці з маркетингу вважають, що цифрова реклама – це зовсім нова область, яка вимагає

нового способу залучення клієнтів та нових способів розуміння того, як клієнти поведуться у порівнянні з традиційним маркетингом.

Цифровий маркетинг орієнтований на конкретний сегмент клієнтської бази та є інтерактивним. Цифровий маркетинг збільшується в масштабах та включає в себе оголошення у результатах пошуку, повідомлення електронної пошти та рекламні твіти – одним словом все, що включає маркетинг з відгуками клієнтів або двосторонню взаємодію між компанією та клієнтом.

Інтернет-маркетинг (також відомий як онлайн-маркетинг) відрізняється від цифрового маркетингу, точніше кажучи є його підмножиною. Інтернет-маркетинг – це реклама виключно в Інтернеті, тоді як цифровий маркетинг може здійснюватися по телефону, на платформах метро, у відеоіграх або через додатки для смартфонів.

У контексті цифрового маркетингу рекламодавці зазвичай називаються «джерелами», а отримувачі цільових оголошень зазвичай називаються «приймачами». Джерела часто використовують високоспеціалізовані, чітко визначені приймачі. Наприклад, після продовження вечірніх годин у багатьох своїх місцях, McDonald's потрібно було розказати про це цільовій аудиторії. Компанія використовувала цільову цифрову рекламу для працівників пізніх змін та мандрівників, оскільки компанія знала, що ці люди складають великий сегмент пізнього нічного бізнесу. McDonald's заохочував їх завантажувати нову програму "Ресторан-шукач", за допомогою цифрової цільової реклами, розміщеної в банкоматах та автозаправних станціях, а також на веб-сайтах, які користувачі часто відвідують ввечері.

Цифровий маркетинг використовується для рекламування не лише продуктів і послуг. Він широко використовується для просування таких речей, як компанії, політичні партії та ідеї. Політичні партії використовують цифровий маркетинг, щоб сповіщати виборців позитивними SMS-повідомленнями про своїх кандидатів та негативними SMS-повідомленнями про опонентів своїх кандидатів, а також «пристосувати» оголошення до одержувачів, які часто використовують певні цифрові канали, такі як новини Facebook та канали

YouTube. McDonald's створила цифрову кампанію "Kick the Trash" для протидії негативній пресі в Німеччині, яка назвала зовнішню частину території компанії брудною.

Цифровий маркетинг створює особливі виклики для своїх постачальників. Цифрові канали швидко розмножуються, і цифрові маркетологи мають слідкувати, як ці канали працюють, як вони використовуються приймачами, і як використовувати ці канали для ефективного маркетингу товарів. Крім того, все важче привертати увагу приймачів, оскільки приймачі все частіше перевантажуються конкуруючими оголошеннями [8]. Цифрові маркетологи також вважають складним аспектом своєї діяльності аналіз величезної кількості даних, які вони фіксують, а потім використовують цю інформацію в нових маркетингових кампаніях.

Проблема збору та ефективного використання даних показує, що для цифрового маркетингу потрібен новий підхід до маркетингу на основі нового розуміння споживчої поведінки. Наприклад, це може вимагати, щоб компанія аналізувала нові форми поведінки споживачів, таких як вподобання на Facebook і твіти у Twitter [9].

Насамкінець, потрібно описати структуру галузі цифрової реклами. Тобто як саме в даній області все влаштовано, навести зв'язки між складовими елементами та їхню взаємодію. На рисунку 1.1 наведено узагальнену структуру ринку цифрової реклами, та наведено основні складові компоненти.

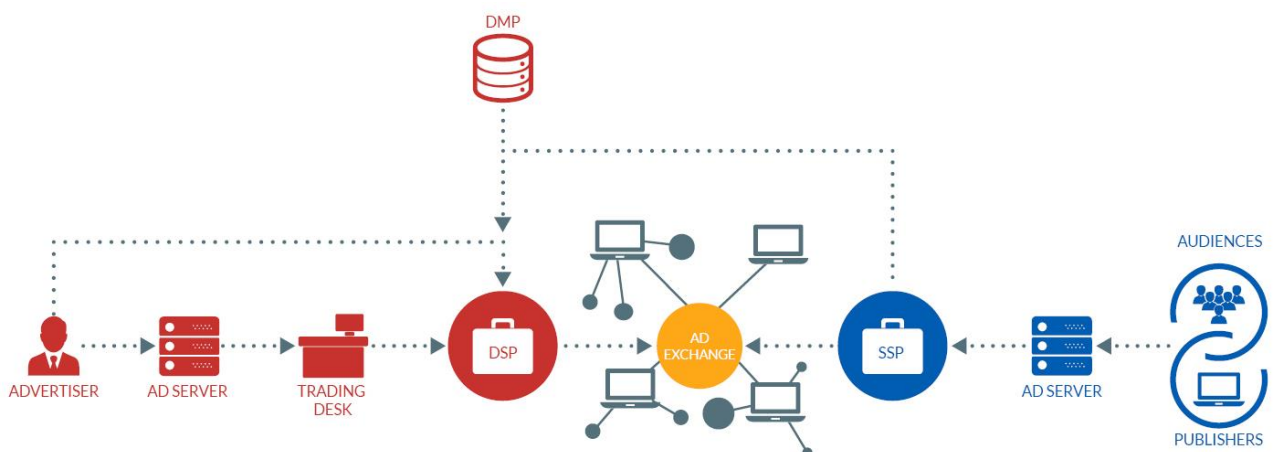


Рисунок 1.1 – Структура ринку цифрової реклами



Важливо розуміти, що у цифровому маркетингу є три основні сторони: рекламодавці, видавці та власне аудиторія, яка переглядає рекламні оголошення. Отже, можна зробити висновок, що саме між цими основним гравцями відбуваються взаємодії в галузі цифрового маркетингу.

Для автоматизації процесів, які відбуваються всередині ринку цифрової реклами, були введені певні проміжні елементи. Наприклад, SSP – програмна платформа, що використовується видавцями для автоматичного продажу свого рекламного інвентаря. У свою чергу, DSP – це так званий аналог платформи SSP для рекламодавців. Вони використовують цю систему для того, щоб купляти рекламні місця на сайтах видавців.

Наведемо більш конкретні визначення складових елементів ринку цифрової реклами, що наведені на рисунку 1.1:

- **Demand-side Platform (DSP)** – програмне забезпечення, яке дозволяє фахівцям із купівлі реклами купувати місця для розташування рекламних оголошень за допомогою ставок на біржах оголошень;
- **Supply-side Platform (SSP)** – програмне забезпечення, яке допомагає видавцям продавати свої рекламні ресурси за найвищою ціною через рекламні біржі;
- **Data Management Platform (DMP)**. Одним із найбільш важливих побічних продуктів зростаючого ринку цифрової реклами – є дані, які маркетологи не знають як використати. Платформа керування даними – це програмне забезпечення, яке сортує та зберігає всі ці дані, фільтруючи їх за потреби таким чином, щоб вони стали корисними для маркетологів та видавців;
- **Trading Desk**. Торгові платформи займаються купівлею та продажем рекламних оголошень від імені своїх клієнтів, використовуючи цифрові технології.

### 1.3 Трансформація цифрової реклами з часом, її типи та види

#### 1.3.1 Трансформація цифрової реклами в часі

За останні 10 років цифровий маркетинг зазнав значних змін. Протягом цього часу люди почали сприймати галузь цифрової реклами перспективною для побудови своєї кар'єри, і багато людей в усьому світі думають, що вона є надійною. Однією з реальних змін у цифровому маркетингу є те, що в цій процедурі автоматизація збільшила об'єми реклами. Окремі особи стають набагато більш залежними від автоматизації цифрового маркетингу. Розширені рекламні кампанії покращили позиції онлайн-бізнесу. Сьогодні люди надзвичайно залежні від електронного бізнесу для їхніх повсякденних потреб.

У наш час Google є лідером серед пошукових систем. Він постійно вдосконалює свій пошук та стає кращим, ніж раніше. Ранні пошукові системи, такі як Yahoo!, AltaVista, Lycos і WebCrawler, зробили критичні кроки для покращення розширеного пошуку. Однак справжні зміни почались в ті роки, коли у 1997 році з'явилась компанія Google. Після появи найпоширенішого інтернет-пошуковика сьогоднішнього дня та вдосконалення його платформ, таких як AdWords у 2000 році, та орієнтації на управління вмістом у 2003 році, в Інтернеті все змінилось.

У 2004 році веб-індекси почали використовувати алгоритми розширеного рейтингу. Перегляд в Інтернеті виявився більш особистим. Ефект, який це мав на бізнес, є величезним. Можливість робити таргетингові програми та створювати системи з урахуванням шаблонів пошуку, призвела до створення радикально нового підходу до спілкування та продажів. Сьогодні пошуковий робот Googlebot виділяє трильйони сторінок в Інтернеті та надає доступ до найбільш відповідних матеріалів за допомогою звичайного або голосового вигляду. Реклама бренду може охопити будь-кого, де завгодно, і це є ефективним [10].

Наведемо перелік 5 найважливіших змін у сфері цифрового маркетингу за останні 10 років:

- а) революція в області рекламних інструментів. Маркетинг та його інструменти змінюються настільки швидко, що з'являється необхідність у відкритті відділу, який би займався дослідженням нових технологій. Ви не можете просто відвідувати маркетингову групу один раз на рік та перебувати в тренді. Ви дійсно потребуєте супроводжуючих осіб, які були б вашими радниками щодо того, якими є основні новинки у галузі цифрового маркетингу на даний момент;
- б) зміни у маркетингових діях. Зміна інструментів маркетингу вказує на необхідність змін у ваших діях. Відео є одним із прикладів. Якщо ви не почали використовувати мобільні телефони та ненав'язливі відеоролики для відображення вашого сайту, то ви не використовуєте наявні сприятливі умови для рекламування, незважаючи на наявність інструментів;
- в) зростання маркетингу в соціальних мережах. Ваш бізнес повинен змінюватися, щоб забезпечити використання соціальних мереж. Але незважаючи на це, не потрібно забувати про те, що потрібно також вдосконалюватись по своїй спеціалізації. Сьогодні ваша компанія повинна бути відкритою книгою. Це зараз дуже важливо, оскільки клієнти хочуть все бачити та контролювати. Сьогодні, за допомогою каналів, таких як YouTube, Instagram, Twitter та Pinterest, з мільярдами клієнтів, прямого маркетингу та комп'ютерних технологій, яким ніколи не потрібен відпочинок, цифровий маркетинг став основою серед ключових частин будь-якої процедури маркетингу;
- г) зміни у маркетинговому бюджеті. У наш час більш гострим стало питання щодо правильного інвестування в маркетинг. Наприклад, бізнес-сторінка у Facebook є безкоштовною. Профіль у Twitter також є безкоштовним. Потрібно вирішити, на які саме додаткові маркетингові заходи потрібно витратити рекламні кошти. Планування маркетингового бюджету стає все більш складним завданням. Деякі грошові ресурси, які раніше вкладались у звичайні ЗМІ, із часом перейдуть до сучасних каналів соціальних медіа;

д) перехід від офлайн до онлайн-маркетингу. Найбільша зміна – це перехід від офлайн до онлайн-маркетингу. Цей процес продовжується завдяки популярності вхідного маркетингу. Ще десять років тому, організації намагались привернути увагу своїх потенційних клієнтів за допомогою повідомлень, причому кінцева мета полягала в тому, щоб бути поміченим. У цей час міра ефективності була критичною. Тобто кількість разів, коли потенційний клієнт побачив комерційну рекламу в значній мірі вирішувала, чи буде клієнт цікавитись тим чи іншим брендом. Наразі те, що роблять організації, полягає у створенні вмісту, який ефективно задовольняє потреби їхніх потенційних клієнтів, щоб залучати їх менш нав'язливо [11]. Саме тому ви бачите організації, які створюють свій блог, подкаст, відео тощо. Вони роблять це не через те, що їм потрібно створювати якийсь новий контент – вони повинні це робити, щоб залишатися у фокусі потенційних клієнтів. У наш час блог є практично основою у процедурах просування в Інтернеті будь-якої організації. У будь-якому випадку, ми повинні пам'ятати одну річ. Вміст, який ви розміщуєте в своєму блозі, не має рекламувати ваші товари чи послуги. Все, що пов'язано з рекламою, може послабити ентузіазм ваших потенційних клієнтів. Ви повинні точно знати час, коли потрібно перетворити ваших потенційних клієнтів на реальних клієнтів.

Тепер наведемо які ще важливі процесі відбувались у цифровому маркетингу з 2007 року:

- пошукова оптимізація зросла. Покращення сайтів для веб-індексування почало стандартизуватися в середині 90-х років, а до 2004 року SEO перетворився на основний рекламний засіб, який використовується у світовому масштабі. Раніше такі процедури, як накладання ключових слів, включення заголовків та створення спам-зв'язків, були запропоновані людьми, які, як правило, не знали нічого кращого. Ніхто не думав про ефект онлайн-соціальних мереж. SEO раніше не був популярним, а

створення шаблонів не було достатньо ефективним. Але все змінилось, і сьогодні SEO є одним із основних засобів у цифровому маркетингу;

- «agile маркетинг» значно зріс. Це є показником того, наскільки продуктивним є створення асоціацій для досягнення рекламних цілей. Рекламна група надає велику кількість методологій та гіпотетичних результатів своїм партнерам з метою швидкого їхнього виконання. До того ж, рентабельність інвестицій є ключовим компонентом, тому що відомі рекламодавці розуміють, що новий бізнес є більш важливим, ніж нові матеріали. Однією з причин зростання гнучкого маркетингу є соціальні мережі. Бренди витратили декілька років, поки зрозуміли, як використовувати ці платформи. Ці канали дали знання про те, що може працювати, і про те, який вміст вони здатні відображати;
- цифровий PR став неймовірно потужним. У недавньому минулому не було такого поняття, як цифрова PR-організація. Були лише звичайні PR організації. Сьогодні все змінилося, оскільки розвиток комп'ютерних технологій не оминув і галузь PR. Ще якийсь час звичайні PR-організації використовували засоби масової інформації для проведення рекламних зустрічей, та представляли своїх клієнтів у пресі, на радіо та на телебаченні. Згодом вони регулярно проводили опитування, щоб накопичувати контактні дані та з'ясовували, чи вони охоплюють свою цільову аудиторію. Сьогодні комп'ютеризовані PR-організації можуть з'ясувати все, що їм треба знати про контакт, перш ніж з ним зв'язуватись. Працюючи з тенденціями в галузі SEO, цифрові PR можуть підкріпити більшість своїх рішень необхідною інформацією, що дозволяє більше зосередити увагу на важливих моментах, та створює більш кількісно виражені результати;
- прозорість. Клієнти в даний час більш швидкі та кмітливі, та здатні ознайомлюватися з новинами в лічені секунди. Клієнти легко порівнюють конкурентів, перевіряють все та розкривають будь-які міфи. Для фахівців із угод, ці зміни демонструють те, що близько 70% потенційних покупців

відсіюється ще до того, як покупець вступить в контакт із подавцями. Потреба в прозорість зрозуміла, коли ви досліджуєте клієнтське враження від брендів. Серед користувачів Інтернету, 19% клієнтів вважають, що легко розрізнити рекламу зв'язків, пропаганду, рекламу брендів та некомерційну рекламу через мережеві мультимедійні засоби, а 30% майже не довіряють вмісту, який вони бачать на Facebook. Дослідження враження покупців стає все більш популярним із кожним днем;

- зміни, які ще не відбулись. Цей аспект зараз важливіший ніж будь-коли раніше. Те що було важливо раніше, стало ще важливішим зараз. Сет Годін склав понад 30 книг, велику кількість записів в блозі та багато разів він обговорював цю тему. Всі ці записи об'єднані однією ідеєю – будьте феноменальними, надзвичайними та унікальними. Сьогодні це означає бути більш значимим, бути кращим у будь-якій справі, якою ви займаєтесь. Цей рекламний стандарт є єдиним, який витримав випробування часом та технологіями. Це хороший аспект для фокусування, коли «земля рухається під ногами».

Варто зазначити, що раніше організації витрачали менше ресурсів, щоб переконати або відмовити клієнтів у покупках. Ділові люди були довіреними, і пряме звернення до організації розглядалося як послідовний крок у здійсненні покупки. Сьогодні клієнти можуть знайти переважну більшість необхідних даних, перш ніж спілкуватися з організацією [12]. Це робить операції та зусилля більш складними для організацій. Сьогоднішній цифровий маркетинг залежить від того, наскільки добре ви розумієте клієнта, щоб ви могли врахувати їхні інтереси та потреби. На сьогоднішній день відомо, що потрібно привертати увагу надзвичайним вмістом, дискусіями, прозорістю, маркетингом через електронну пошту, маркетингом через чутки та вірусною рекламою.

### 1.3.2 Типи цифрової реклами

"Цифровий маркетинг" – це один з тих термінів, який ми всі відчуваємо, що знаємо, але коли справа доходить до того, що потрібно його реально використати, ми чомусь застрягаємо. Зрештою, не зовсім зрозуміло, про які типи цифрового маркетингу ми говоримо. Тому замість того, щоб сказати щось дуже загальне, на кшталт "це пов'язано з онлайн-маркетингом" або "загальний термін для всіх онлайн-маркетингових заходів", можна просто сказати наступне. Цифровий маркетинг – це всі маркетингові заходи, що використовують цифрові пристрої [13].

Це включає в себе інтернет-маркетингову діяльність – але, строго кажучи, це також включає в себе деякі офлайн-маркетингові заходи. І це добре. Чому? Оскільки було б помилкою розглянути дії онлайн-маркетингу, повністю відірвані від усіх інших маркетингових заходів.

Для того, щоб дати глибоке розуміння того, що включає термін "цифровий маркетинг", ось десять найважливіших типів цифрового маркетингу:

1. Маркетинг в соціальних мережах. Останнім часом соціальні мережі набувають все більшої популярності. Тому із них ми почнемо наш список типів цифрового маркетингу. Соціальні медіа, звичайно, мають законно зароблене місце в цьому списку. Соціальний медіа-маркетинг – це "використання платформ та веб-сайтів соціальних мереж для просування продукту чи послуги". Це означає, що весь обмін інформацією та взаємодія з послідовниками, шанувальниками, партнерами чи конкурентами на платформах соціальних мереж з метою просування вашої продукції є частиною цифрового маркетингу. По суті, маркетинг в соціальних мережах – це цілеспрямоване використання розмов у соціальних мережах з метою підвищення обізнаності про бренд чи продукт;
2. Контент-маркетинг. Контент-маркетинг – це мистецтво використання розповіді та цінної інформації для підвищення рівня поінформованості про бренд з метою досягнення вигідної дії від вашої цільової аудиторії.

Контент-маркетинг спрямований на побудову відносин з потенційними клієнтами, і висвітлення компанії в ролі партнера, а не рекламодавця. Контент-маркетинг – це стратегічний маркетинговий підхід, орієнтований на створення та поширення цінного, релевантного та послідовного вмісту для залучення та збереження чітко визначеної аудиторії – і, в кінцевому рахунку, залучення прибуткових дій клієнта. Оскільки контент-маркетинг сильно залежить від поширення вмісту, контент-маркетинг рідко можна розглядати як повністю відокремлений від інших типів цифрового маркетингу, які можуть заповнити частину розповсюдження вмісту;

3. Пошукова оптимізація (SEO). SEO – це процес оптимізації вмісту чи веб-сайтів, щоб вони відображалися в результатах пошуку в пошукових системах, наприклад, в Google. Пошукові системи вирішують, які веб-сайти повинні показуватись за пошуковим запитом на основі ключових слів, згаданих на веб-сайті, і посилань, які стосуються цього веб-сайту. Це означає, що SEO значною мірою полягає у використанні правильних ключових слів або ключових фраз у копії веб-сайту або вмісту, який ви хочете показати під час пошуку та отримання посилань на цей веб-сайт або вміст. Пошукова оптимізація (SEO) – це процес оптимізації вашого онлайн-контенту, так що пошуковий механізм буде показувати його як найпопулярніший результат для пошуків певного ключового слова. Існує кілька стратегій для оптимізації веб-сайтів для пошукових систем – деякі з них приймаються компанією Google, а інші вважаються досить сумнівними, і Google може прийняти рішення щодо покарання вашого веб-сайту. Існує тісний зв'язок між контент-маркетингом, маркетингом в соціальних мережах та SEO. Найбільш очевидним твердженням є те, що соціальні медіа та SEO є ідеальними каналами для розповсюдження контенту;
4. Маркетинг в пошукових системах (SEM). Якщо SEO описує процес отримання безкоштовного трафіку з пошукових систем, то SEM відноситься до платного трафіку з пошукових систем. Найпоширенішою



формою маркетингу в пошукових системах, мабуть, є Google Adwords з тієї простої причини, що Google на сьогодні є найпопулярнішою пошуковою системою. У SEM маркетинг, як правило, сплачує пошуковій системі заздалегідь визначену суму грошей для показу маркетингового повідомлення в різних місцях в результатах пошуку за певними ключовими словами або фразами. SEM – це форма реклами «Pay-Per-Click»;



Рисунок 1.2 – Маркетинг в пошукових системах

5. Pay-Per-Click (PPC). Подібно до SEM, інші форми реклами «Pay-Per-Click» також описують маркетингові методи, за якими маркетолог платить за кожен клік на посилання на веб-сайт. Крім пошукових систем, практично всі соціальні мережі пропонують можливість оформлення реклами «Pay-Per-Click». Ця реклами з'являється в каналі цільових користувачів соціальної мережі. Варто зазначити, що багато із зазначених вище дій не можна чітко розділити. Межі між різними типами цифрового маркетингу часто розмиті;
6. Партнерський маркетинг. Партнерський маркетинг – це типовий цифровий маркетинг, який націлений на результат. На відміну від реклами «Pay-Per-Click», при партнерському маркетингу, рекламодавець не платить за трафік, а замість цього платить за конверсію. Розцінки, як правило, вищі, але ризик для рекламодавця мінімальний, оскільки він сплачує лише за конверсію. Партнерський маркетинг користується популярністю у

блогерів та власників веб-сайтів з високим рівнем трафіку, які заробляють гроші, продаючи продукцію інших людей своїй аудиторії;

7. E-mail маркетинг. Маркетинг із використанням електронної пошти – це один з найкращих конверсійних маркетингових каналів. Надсилаючи регулярні оновлення для ваших абонентів через електронну пошту, ви можете створювати та розвивати відносини. Надсилаючи корисні новини за допомогою електронної пошти, ви можете створити довіру – врешті-решт, ви зможете перетворити певний відсоток вашої аудиторії на клієнтів. Варто зазначити, що маркетинг із використанням електронної пошти це набагато більше, ніж купівля списків електронної пошти, та розсилання усім хто є в цьому списку рекламних повідомлень або одноразове рекламне оголошення в чужій електронній інформаційній розсилці. Найкращі результати у маркетингу із використанням електронної пошти завжди будуть досягнуті за допомогою списку підписників, яких ви особисто залучали, та з якими вибудовували партнерські стосунки. Тоді ваш список електронних адрес може легко перетворитися на ваш кращий цифровий маркетинговий актив;
8. Багато статей про цифровий маркетинг зупиняються на вищезгаданих 7 типах онлайн-цифрового маркетингу. Але тоді ми могли б просто називати це онлайн-маркетингом. Але також варто знати про найважливіші типи офлайн-цифрового маркетингу. Перший тип – це радіо-реклама. Це саме ті перерви у вашій улюбленій музичній програмі, де хтось каже вам, яку машину купити або який очищувач використовувати. Деякі з цих місць можуть бути досить дратівливими, деякі з них можуть розважати – принаймні, коли ви їх чуєте вперше. Існує ще більше радіо реклами: наприклад, ви можете спонсорувати радіо-програму, в якій буде згадуватись ім'я вашого бренду;
9. Телевізійна реклама. Вона ще не мертва: це та пауза, яку ви бачите при перегляді вашої улюбленої телевізійної програми. Попри те, що ми все більше вдаємось до пошуку способів уникнути перегляду телевізійних

реklamних роликів, у телевізійній рекламі все ще є певні маркетингові успіхи. Насправді, деякі рекламні оголошення, подібні до створених для «Super Bowl», роблять телевізійну рекламу досить успішною. Деякі з найкращих рекламних оголошень для «Super Bowl» навіть набувають значної популярності в соціальних мережах, і дозволяють привернути частину клієнтів, за яких рекламодавцям не треба платити – або, можливо, цей аспект вже включений в ціну рекламного простору «Super Bowl». Але потрібно мати на увазі те, що реклама через телебачення пропонує дуже обмежені можливості таргетування;

10. Реклама на мобільних телефонах. Звичайно, багато з вищезгаданих типів цифрового маркетингу можуть мати місце на вашому мобільному телефоні. Ви, ймовірно, використовуєте Facebook, Instagram і Google на своєму мобільному телефоні. Але на вашому мобільному телефоні є й інші види маркетингу, які до цих пір не вписуються в жодну з названих категорій. Наприклад, SMS-реклама вже існує давно. Із такими технологіями, як Bluetooth, який ще досі розвивається, варіанти для маркетингу та реклами на мобільних пристроях, особливо в місцевому маркетингу, на сьогоднішній день ще не були повністю використані.

Варто зазначити, що можна далі сперечатись із приводу того, що ще належить до «цифрового маркетингу». Проте для побудови власної цифрової маркетингової стратегії, ви повинні знати про ці 10 – і вибрати ті, які, на вашу думку, найкраще підходять для ваших цілей. Потрібно пам'ятати, що всі вони взаємодіють один з одним, і часто межі між ними є досить розмитими [14].

### 1.3.3 Види цифрової реклами

Останнім часом цифрова реклама набуває все більшої популярності. Серед цифрової реклами на даний момент найбільше виділяється онлайн-реклама, або

її ще часто називають Інтернет-рекламою. Саме тому в даній роботі Інтернет-рекламі буде приділена основна увага.

При безпосередньому розгляді Інтернет-реклами виникає логічне запитання: які існують формати рекламних оголошень для проведення маркетингових кампаній? Якщо дивитися правді в очі, то потрібно сказати, що немає такого поняття як універсальний формат оголошення [15]. Саме тому давайте розглянемо доступні види рекламних оголошень – від традиційних до нових, які тільки набирають популярність.

Отже, розглянемо основні види рекламних оголошень, які існують на даний момент:

- стандартний банер. Банерні оголошення бувають різних розмірів; статичні або анімовані, зазвичай займають верхню, нижню або бокову частину екрана. Це один із найбільш популярних рекламних форматів на сьогоднішній день. Раніше ці оголошення виробляли в основному неефективні, випадкові кліки. Тепер вони можуть бути досить ефективними, якщо реалізуються в належному контексті. Банер пасивний, не нав'язливий і не перериває роботу користувача. На жаль, через цю причину виникає так звана "банера сліпота" (ситуація, коли в користувачів виникає несвідомий опір до банерів). Успіх банерної реклами значною мірою залежить від відомості бренда. Банерні реклами універсальні та прості. Ви можете їх швидко створювати та розміщувати на сайті. За словами різних фахівців, дизайнер банерів потребує ретельного балансу із тонким привабливим вмістом, але не настільки щоб відволікати користувача від його справ.

Плюси:

- 1) доступний на всіх екранах;
- 2) може бути запущений у великих обсягах;
- 3) найдешевший формат;
- 4) дуже швидко розгорнути;
- 5) легко інтегрувати.

Мінуси:

- 1) обмежений простір;
  - 2) швидко потрапляють у синдром "банерної сліпоти";
  - 3) Незважаючи на боротьбу з відсутністю простору та синдромом "банерної сліпоти" користувачів за допомогою анімованого вмісту, банери не настільки ефективні, як інші більш просунуті формати.
- міжсторінкова реклама. Проміжні оголошення – це зображення, які охоплюють весь розмір екрану. Це дозволяє викликати більше зацікавленості, а отже, підняти ваш показник «SAR» для кращого просування бренду та більшого відгуку. Вони, як правило, відображаються в найважливіші моменти під час навігації, як-от відкриття, перегортання сторінок або в іграх між рівнями. Це вимагає дії користувача; зазвичай кліку на кнопку, щоб закрити оголошення, та перейти до потрібного вмісту. Рекламодавці можуть створювати чудовий та цікавий вміст із високоякісними та переконливими оголошеннями [16]. Це може зменшити незадоволення в користувачів. На думку багатьох експертів, проміжні оголошення найкраще розміщувати в іграх із рівнями. Природні паузи в іграх між рівнями дозволяють міжсторінковим оголошенням бути ненав'язливими. Рекомендується не переривати взаємодію з користувачем. Тому видавці не повинні розміщувати проміжну рекламу посередині гри.

Плюси:

- 1) більший розмір;
- 2) ширше повідомлення, експозиція та швидкість запам'ятовування;
- 3) візуально привабливий;
- 4) велика кількість показів;
- 5) висока конверсія;
- 6) може містити анімований та інтерактивний мультимедійний контент.

Мінуси:

- 1) може бути дуже нав'язливим – у разі поганого розміщення;

- 2) високий рівень параметра «CTR» може виникнути через труднощі з закриттям міжсторінкового оголошення;
  - 3) потрібно більше дизайнерських робіт і роздумів щодо правильного розміщення.
- Pre-roll – рекламний ролик, який програватиметься на сторінках сайтів, де є відео-контент, перед відео в тому ж плеєрі. Досить яскравий приклад, це сайт сервісу «Megogo». У даному випадку рекламний ролик програватиметься перед початком фільму.

Плюси:

- 1) найбільш якісний, гарний та правильний відео-інвентар для рекламодавців;
- 2) велика кількість переглядів.

Мінуси:

- 1) мало трафіку для купівлі;
  - 2) у світі мало сайтів із вбудованим відео-контентом. Більше того, там де є відео-контент – в більшості випадків, це тільки розділ на сайті;
  - 3) так як мало трафіку, то важко його купити, тому конкуренція та ціна на такий інвентар є високою.
- Content-roll – рекламний ролик, який показується на початку, всередині або в кінці статті чи будь-якої іншої сторінки сайту, коли користувач гортає сторінку сайту вниз. Після того, як відео закінчиться плеєр зникає.

Плюси:

- 1) content-roll можна поставити на будь-який сайт. Не потрібно мати відео-контент на сайті, content-roll можна поставити в будь-якій статті чи сторінці;
- 2) більша кількість трафіка в порівнянні з форматом pre-roll.

Мінуси:

- 1) невелика кількість виконаних показів. Користувачі часто швидко гортають сторінку вниз, через що content-roll або не встигає

відкритись, або рекламне відео відтворювалось менше 2-х секунд в полі зору користувача щоб рекламний показ був зарахований;

- 2) загалом і в цілому підходить тільки для дуже великих видавців. На маленькому об'ємі трафіка на сайті, даний формат не дуже ефективний, так як генерує занадто малу кількість показів, щоб вони перетворились в значну суму доходу для видавця [17]. В першу чергу content-roll на маленьких видавцях не ефективний і для самих компаній, що займаються рекламою, оскільки з точки зору витрат на налаштування і кінцевого результату у вигляді доходу він себе не окуповує.

- video-banner – це рекламне відео, яке програвється в стандартному банері, зазвичай розміром 300 \* 250 пікселів.

Плюси:

- 1) багато трафіку. Якщо банер знаходиться на першому екрані, то практично відео-банер здатний генерувати таку ж кількість відео трафіку як і при звичайних показах банера;
- 2) набагато більша кількість відео трафіка в порівнянні з форматами pre-roll і content-roll (За умови що банер не знаходиться в шапці сторінки).

Мінуси:

- 1) більшість рекламодавців не використовує формат відео-банера;
- 2) досить часто відео-банери використовують для шахрайського трафіка. Так як фактично показ відео-банера продають як показ формату pre-roll, що є абсолютно різними речами з точки зору залучення користувача у відео-контент. У даному випадку це є ключовим недоліком для рекламодавців. З 2017 року боротьба з шахрайськими відео-банерами вийшла на новий рівень. Однак, для відео-платформ, SSP, мереж відео-реклами банер є одним з найбільш прибуткових форматів, яким не гребує навіть компанія Google, яка є лідером в галузі цифрової реклами [18].

- native video-roll – рекламний формат, який по суті є поєднанням форматів content-roll та native-banner в одному блоці. Розміщується в кінці статей, де зазвичай знаходяться нативні блоки. Основна ідея полягає в тому, щоб привернути більше уваги ніж нативні банери за рахунок додавання в блок формату «content-roll».

Плюси:

- 1) аналогічні до формату «content-roll»;
- 2) можливість для рекламної компанії встановлення банерів по собівартості, але при цьому заробляти на форматі «content-roll».

Мінуси:

- 1) на практиці цей формат не відразу підходить видавцям. Як правило, справа доходить до такого формату з тими видавцями, з якими вже успішно випробуваний якийсь формат, і їм просто пропонується його розширити до нативного відео-блоку;
  - 2) при додаванні в нативні блоки, які знаходяться на сайті, цього формату, CPM по нативним блокам падає, так як відео-плеєр автоматично при розгортанні зміщує їх вниз, що знижує CTR по банерах.
- In-game video-roll – рекламний формат, який являє собою рекламні ролики всередині ігор. Зазвичай мають на увазі онлайн-ігри.

Плюси:

- 1) унікальний відео-інвентар для рекламодавців;
- 2) низький поріг конкуренції, не всі компанії надають даний сервіс;
- 3) Підтримка HTML та флеш ігор.

Мінуси:

- 1) вузька ніша (тільки онлайн-ігри);
- 2) висока залежність від аудиторії онлайн-ігор. Наприклад CPM по онлайн-іграх в певному регіоні може коливатись від \$0.4 на низькоякісних сайтах до \$5 на популярних;



- 3) відсутня універсальна методологія налаштування реклами. В 90% випадків команді «Ad Ops» потрібно робити індивідуальні налаштування для кожного видавця.
- top-roll – рекламний відео-формат, для якого фактично немає встановленої назви на ринку. Кожен, хто пропонує такий формат, називає його по-своєму. Досить поширеною є назва «top-roll», так як він показується в шапці сайту.

Плюси:

- 1) знаходиться відразу в зоні видимості користувача. Поки користувач знаходиться на першому екрані, і думає що йому далі робити на сайті – в цей час вже відправляється запит на відео-показ;
- 2) досить велика кількість відео-показів. Значно перевершує формат «content-roll» по таким параметрам як запити та покази;
- 3) високий показник «Active View»;
- 4) низька конкуренція.

Мінуси:

- 1) формат відноситься до категорії «Rich Media», тобто не стандартний формат, на який необхідно «вмовити» видавця.
- video-slider – відносно новий відео-формат, який з'явився кілька років тому. Його поява пов'язана з великою кількістю запитів на відео-інвентар на ринку. Video-slider – це рекламний формат, який зазвичай з'являється в правому нижньому кутку екрану. Даний формат можна закрити або згорнути – через 5 секунд з'являється відповідний хрестик.

Плюси:

- 1) генерує в 2 рази більше відео-показів ніж формат «content-roll»;
- 2) генерує багато відео-трафіка;
- 3) можна поставити на всіх сторінках сайту;
- 4) автовідтворення без звуку. Звук з'являється тільки при наведенні;
- 5) не є шахрайським форматом, на відміну від відео-банера.

Мінуси:

- 1) при всіх своїх перевагах, video-slider – це «Rich Media» формат, тому видавця треба додатково «вмовляти»;
- 2) на ринку відсутня певна стандартизована метрика ефективності для формату «video-slider».

#### 1.4 Огляд математичних методів для моделювання і прогнозування характеристик трафіка цифрової реклами

Прогнозування є ключовою частиною процесу створення маркетингових планів як для річних інвестицій, так і для окремих кампаній. Створення більш точних прогнозів щодо прибутків від цифрового маркетингу зараз має важливе значення враховуючи зростання інвестицій у цифровий маркетинг. Тому для сучасного маркетолога важливо навчитися використовувати дані та аналітичні інструменти для отримання прогнозів з достатньо точністю.

Дані змінили маркетинг. Ніколи раніше маркетинг не був таким вимірюваним. Із вузькоспеціалізованої діяльності в науково-дослідних підрозділах великих організацій, маркетингова аналітика стала звичною справою в більшості рекламних компаній. Вимірювання ефективності маркетингу все більше визнається ключем до маркетингових успіхів.

На відміну від продажів, маркетинг повинен розглядатись в довгостроковій перспективі. У продажах керуються швидкими рішеннями, що відповідають змінам на ринку. На відміну від цього, маркетинг формується за довгостроковими тенденціями. Тенденції, які не були завчасно визначені, можуть в кращому випадку вплинути на ефективність маркетингової кампанії. А в гіршому випадку, завдати шкоди конкурентоспроможності або навіть життєздатності бізнесу.

Саме тому з'являється необхідність створення маркетингових прогнозів. В ідеальному світі ваша організація повинна мати культуру, яка поважає дані та аналітики. На жаль, лише кілька організацій структуровані таким чином. Це

неправильно, оскільки переваги розробки маркетингової стратегії, що заснована на даних, є суттєвими, в тому числі:

- підвищення ролі відділу маркетингу. Зосередження уваги на результатах маркетингу (підвищення обізнаності, генерації запитів, потенційних клієнтів та продажів), а не вхідних даних (бюджетів, людей, ресурсів, часу) встановлює маркетинг як центр доходів, який підлягає підтримці, а не центр витрат, який потрібно зменшити;
- більше обґрунтування витрат на маркетинг – маркетинг, що заснований на даних та прогнозування допомагають створити бізнес-кейси для маркетингових витрат. Професійні, фактично підтверджені прогнози демонструють вплив рекламних заходів на мінімальний рівень доходів;
- більша відповідальність – коли керівники відділу маркетингу беруть на себе більшу пряму відповідальність за доходи, їхні зарплата починає виглядати більше схожою до їхніх колег з відділу продажів. Не кожен СМО почуває себе комфортно при системі зарплати, яка залежить від доходу. Однак, якщо маркетологи почнуть відслідковувати та вимірювати все, що вони роблять, то, швидше за все, вони почнуть приділяти більше уваги до маркетингових цілей;
- стратегічні переваги бізнесу. Бізнес, який може краще передбачити майбутнє, має кращі можливості для процвітання в ньому. Історія наводить нам багато прикладів організацій, які не зрозуміли, як змінюється навколишній світ. Але не прогнозування є основною проблемою, організаціям заважає менталітет «бізнес як завжди». Це робить їх невідповідними до майбутнього та вразливими для атаки з боку конкурентів;
- перебування на сучасному рівні – цифровий маркетинг постійно змінюється. Те, що сьогодні є провідним, швидко стане застарілим завтра. Прогнозуючи майбутнє, маркетологи менш імовірно пропустять «наступну значущу річ» та залишаться позаду.

Постійні зміни змушують організації завжди думати наперед і планувати частіше. Маркетингова аналітика раніше виконувала лише роль допоміжного інструменту. Але зараз вона стала одним із основних засобів при здійсненні стратегічно планування.

Варто зазначити, що рекламний трафік, який генерується, має велику кількість метрик та параметрів. Зазвичай статистичні дані щодо трафіка цифрової реклами може містити 20 та більше різноманітних характеристик. Для того щоб спростити аналіз цього набору даних, інколи доцільно виділяти так звані ключові характеристики трафіка. Досить часто цими ключовими характеристиками є параметри доходу та заповнюваності («fill rate» – співвідношення кількості показів реклами до загальної кількості запитів). Важливість доходу не викликає сумнівів, оскільки визначає успішність бізнесу як рекламодавців, так і видавців. Звідси випливає, що він впливає також на успішність рекламних компаній-посередників. Другою важливою характеристикою є показник заповнюваності. Він є насамперед важливим для видавців, оскільки характеризує чи переглядають рекламу відвідувачі сайту. Але оскільки від цього показника залежать інші характеристики, зокрема дохід, то він автоматично стає важливим для всіх учасників бізнесу в галузі цифрової реклами.

Як вже було зазначено вище, рекламний трафік цифрової реклами характеризується набором метрик, тобто сукупністю числових показників. Причому ці показники фіксуються через однакові проміжки часу. Тобто, можна зробити висновок, що ми маємо справу із часовими рядами цих характеристик. Ці висновки значно спрощують задачу прогнозування характеристик трафіка цифрової реклами, оскільки ми принаймні знаємо, що ми маємо справу із часовим рядом. А для них на даний момент розроблена велика кількість різноманітних методів побудови моделей для прогнозування.

Таким чином, ми прийшли до висновку, що математичним апаратом, який слід використати для побудови моделей прогнозу характеристик трафіка

цифрової реклами, є часові ряди. Зокрема потрібно використати прогностичні моделі часових рядів.

Найбільш поширеними моделями для прогнозування даних у вигляді часових рядів є такі моделі як:

- модель авторегресії;
- модель авторегресії та ковзного середнього;
- модель авторегресії та інтегрованого ковзного середнього;
- модель авторегресії та інтегрованого ковзного середнього з екзогенними змінними.

#### Висновки до розділу 1 та постановка задачі дослідження

У даному розділі було розглянуто актуальність прогнозування характеристик трафіка цифрової реклами. Зокрема, наведено опис галузі цифрової реклами та показано, що дана галузь зараз стрімко розвивається. При цьому зазначено, що прогнозування характеристик трафіка цифрової реклами дає змогу побачити їхні майбутні значення, що є досить корисною інформацією при здійсненні планування.

Також описане поняття цифрової реклами – сукупності всіх маркетингових заходів, що використовують цифрові пристрої. До того ж наведені основні підходи до здійснення цифрового маркетингу та його основні задачі. А ще наведено опис структури галузі цифрової реклами, тобто наведено схему, яка пояснює «як все працює» в області цифрового маркетингу.

Після цього було описано трансформацію цифрової реклами в часі, показано як змінювалась форма подання реклами. А також наведені основні типи та види цифрового маркетингу на даний момент.

Насамкінець було наведено огляд математичних методів для моделювання та прогнозування характеристик трафіка цифрової реклами. Описано чому в галузі цифрового маркетингу такі важливі прогнози, та наведені основні математичні моделі для побудови прогнозів.

Основною задачею магістерської дисертації є дослідження моделей та підходів до прогнозування характеристик трафіка цифрової реклами. Зокрема, побудова моделей для прогнозування деяких ключових характеристик рекламного трафіка, а також подальше порівняння побудованих моделей та вибір кращої. Серед показників, які будуть досліджуватись обрані наступні дві характеристики: дохід та показник «fill rate» (співвідношення кількості показів реклами до загальної кількості запитів).

Щоб розв'язати поставлену задачу, потрібно здійснити наступні кроки:

- а) виконати загальний огляд моделей часових рядів та вибрати ті, які відповідають поставленій задачі прогнозування характеристик трафіка цифрової реклами;
- б) побудувати моделі часових рядів за допомогою обраних підходів з використанням статистичних даних;
- в) порівняти отримані моделі та вибрати найкращу;
- г) виробити рекомендації щодо вдосконалення у майбутньому розроблених підходів до розв'язання даної задачі.

## РОЗДІЛ 2 МЕТОДИКА МОДЕЛЮВАННЯ ТА ПРОГНОЗУВАННЯ ХАРАКТЕРИСТИК ТРАФІКА ЦИФРОВОЇ РЕКЛАМИ

### 2.1 Аналіз та конкретизація поставлених задач

На початку дослідження була поставлена задача прогнозування деяких ключових характеристик трафіка цифрової реклами. А саме – доходу та показника «заповнюваності». Проаналізуємо поставлену задачу та опишемо її більш детально. По суті нам потрібно виконати наступне завдання. На основі статистичних даних, які збираються самою рекламною компанією, потрібно зробити середньостроковий або довгостроковий прогноз показників доходу та заповнюваності.

Оскільки потрібно зробити прогноз на основі статистичних даних, які збирались із однаковим інтервалом, то маємо справу із часовими рядами. Саме тому, в даному розділі будуть розглянуті моделі часових рядів та обрані ті з них, які відповідають поставленій задачі прогнозування даних. Найбільш популярними моделями прогнозування часових рядів є моделі авторегресії, авторегресії з ковзним середнім та похідні від цих моделей. Тому зосередимо увагу саме на цих моделях.

Після вибору конкретних методів розв’язання поставленої задачі, потрібно безпосередньо провести прогнозування обраних характеристик. Побудова бажаних моделей буде здійснюватися шляхом їхнього поступового ускладнення, тобто будемо рухатись від простого до складного.

Наступним кроком буде порівняння побудованих моделей. Порівнювати потрібно з точки зору точності прогнозу, а також із використанням різних інформаційних критеріїв. Після того, як буде здійснене порівняння моделей за точністю, потрібно вибрати найкращу модель з них. Як тільки найкраща модель буде вибрана, можна буде оцінити значення доходу та заповнюваності в недалекому майбутньому.

Після того, як наведені вище процедури будуть виконані, потрібно перейти до заключного етапу, а саме – виробити рекомендації щодо вдосконалення у

майбутньому розроблених підходів до розв'язання даної задачі, для того щоб покращувати розроблені моделі відповідно до мінливого ринку цифрової реклами.

## 2.2 Методика побудови математичних моделей на основі статистичних даних у формі часових рядів

«Час» – це найважливіший фактор, що забезпечує успіх у бізнесі. Інколи буває важко йти в ногу з часом. Але, завдяки технологічному прогресу, було розроблено декілька потужних методологій, за допомогою яких, в певному сенсі, ми можемо "бачити речі" наперед. Зокрема мова йде про методи передбачення та прогнозування. Одним з таких методів, який працює з даними на основі часу, є моделювання часових рядів. Як впливає з назви, це включає в себе роботу з даними на основі часу (років, днів, годин, хвилин), для того, щоб отримати приховані закономірності для прийняття обґрунтованих рішень [19].

Моделі часових рядів є дуже корисними, якщо у вас є послідовні дані, що містять певні закономірності. Більшість ділових будинків працюють із часовими рядами, для того, щоб аналізувати кількість продажів на наступний рік, трафік веб-сайту, позиції конкурентів та багато іншого.

Перед тим, як перейти до опису прогностичних моделей, спочатку наведемо базові терміни та передумови для використання бажаних моделей.

Розпочнемо з поняття **стаціонарності**. Існує три основних критерії, які забезпечують стаціонарність часового ряду:

- а) Середнє значення ряду не повинне бути функцією від часу, а навпаки має бути постійним. На рисунку 2.1, графік зліва задовольняє умови стаціонарності з точки зору сталого середнього значення часового ряду, тоді як часовий ряд на графіку справа, що має червоний колір, має середнє значення, яке залежить від часу;



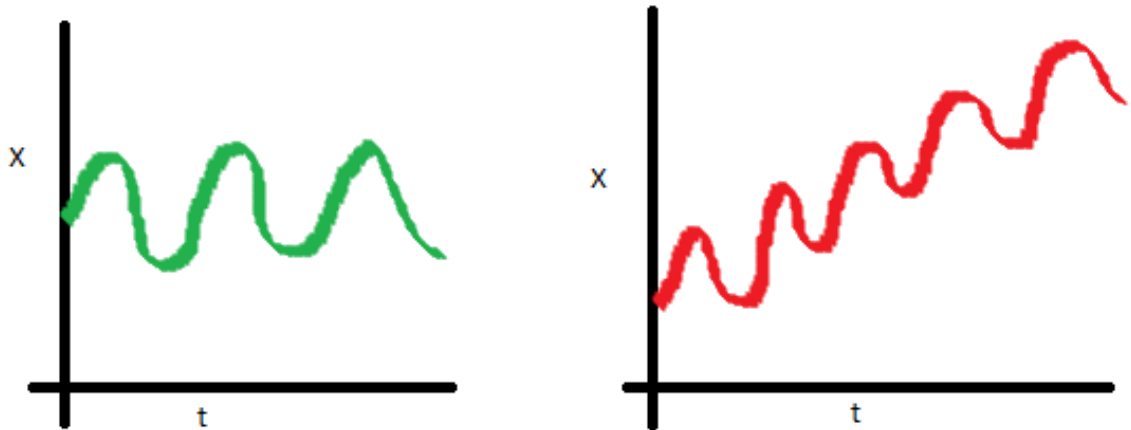


Рисунок 2.1 – Постійне та змінне середнє значення часового ряду відповідно

б) Дисперсія часового ряду не повинна бути функцією від часу. Ця властивість називається «гомоскедастичність». Рисунок 2.2 показує, який часовий ряд є стаціонарним з точки зору дисперсії, а який – ні. У даному випадку потрібно звернути увагу на різний розмах розподілу на графіку праворуч;

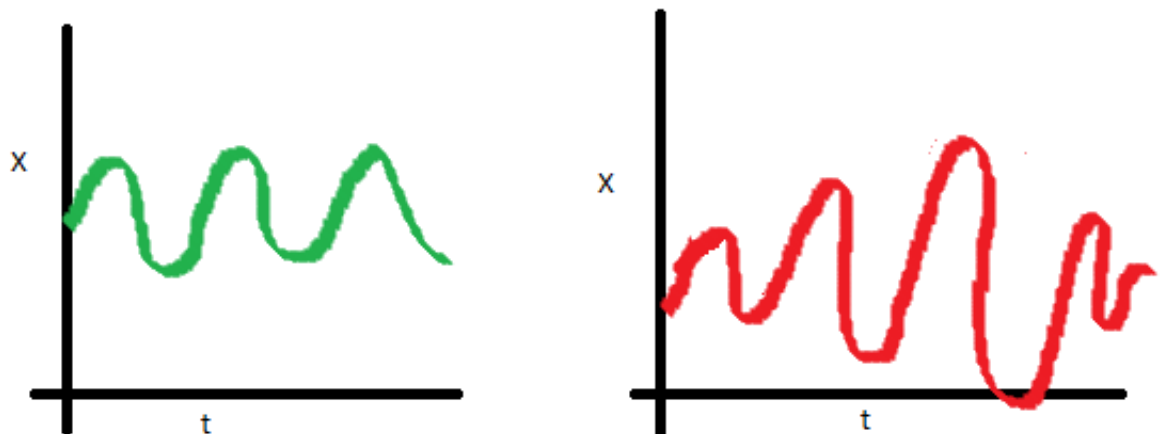


Рисунок 2.2 – Постійна та змінна дисперсія часового ряду відповідно

в) Коваріація часового ряду в  $i$  – й момент часу та  $(i + m)$  – й момент часу не повинна бути функцією від часу. На рисунку 2.3, на графіку праворуч, можна побачити, що значення часового ряду знаходяться все ближче одне до одного зі збільшенням часу. Отже, в даному випадку коваріація не є константою, і змінюється з часом для «червоного часового ряду». Тому маємо стаціонарний часовий ряд, з точки зору коваріації, ліворуч, та нестаціонарний – праворуч.

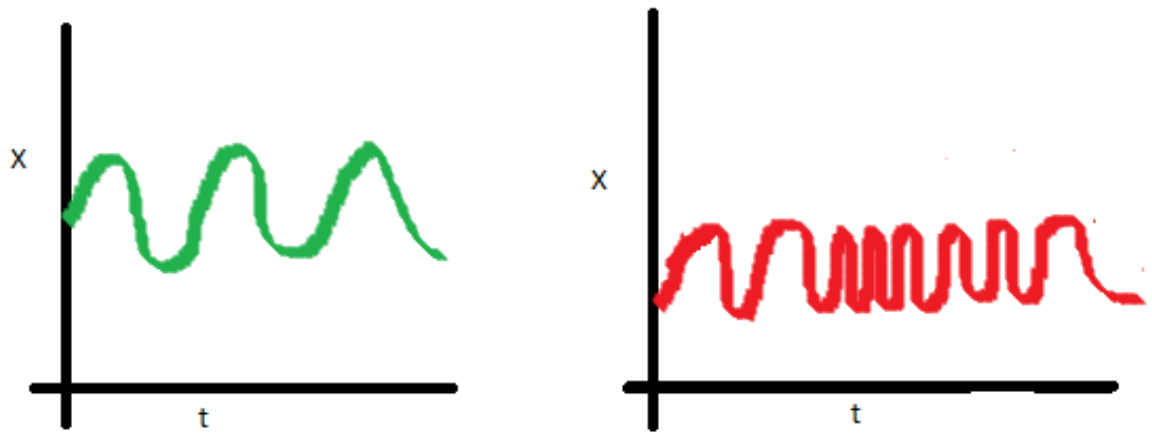


Рисунок 2.3 – Постійна та змінна коваріація часового ряду відповідно

Тепер виникає природне запитання – для чого взагалі досліджувати часовий ряд на стаціонарність. Власне причина, через яку це дослідження потрібно проводити на початку, полягає в тому, що доки ваш часовий ряд не стане стаціонарним, ви не зможете побудувати для нього бажані моделі. У випадках, коли критерій стаціонарності порушується, першим кроком має стати процедура приведення часового ряду до стаціонарного вигляду, а вже потім можна застосовувати стохастичні моделі для прогнозування цього часового ряду [20]. Існує кілька способів приведення часових рядів до стаціонарного вигляду. Наведемо деякі з них: видалення тренду, застосування різниць і т. д.

Наступним важливим поняттям є **«випадкове блукання»**. Це найважливіша концепція часових рядів. Більшість спеціалістів добре знати цю концепцію. Але в галузі інформаційних технологій є досить багато людей, які інтерпретують випадкові блукання як стаціонарний процес. Проте насправді це не так, і в даному розділі буде наведене обґрунтування даного твердження. Давайте приведемо приклад: уявіть собі дівчину, яка рухається випадковим чином на гігантській шаховій дошці. У цьому випадку наступна позиція дівчини залежить тільки від її останньої позиції. Графічне представлення даного прикладу наведено на рисунку 2.4.



Рисунок 2.4 – Приклад «випадкового блукання»

Тепер уявіть, що ви сидите в іншій кімнаті і не можете бачити дівчину. При цьому ви хочете прогнозувати позицію дівчини залежно від часу. Наскільки точним буде такий прогноз? Ясна річ, що чим більше кроків зробить дівчина, тим більше прогноз стає неточним. При  $t = 0$  ви точно знаєте, де знаходиться дівчина. На наступному кроці вона може переміститися лише до 8 квадратів, і тому ймовірність того, що ваше передбачення вірне, знизиться до  $1/8$  замість 1 в попередньому випадку. При цьому дана ймовірність буде продовжувати знижуватися. Тепер наведемо часовий ряд, який описує цю ситуацію:

$$X(t) = X(t - 1) + \varepsilon(t) \quad (2.1)$$

де  $\varepsilon(t)$  – це похибка в момент часу  $t$ . Вона характеризує ту випадковість, яку вносить дівчина в кожен момент часу.

Тепер, якщо ми рекурсивно розпишемо всі значення  $X(s)$ , то в кінцевому результаті ми дійдемо до такого рівняння:

$$X(t) = X(0) + \sum_{i=1}^t \varepsilon(i) \quad (2.2)$$

Тепер давайте спробуємо перевірити наші припущення щодо стаціонарності для цього формулювання випадкового блукання:

а) Чи є математичне сподівання даного часового ряду константою?

$$E[X(t)] = E[X(0)] + \sum_{i=1}^t E[\varepsilon(i)] \quad (2.3)$$

Як відомо, математичне сподівання будь-якої похибки є нульовим, оскільки вона випадкова.

Тоді маємо наступне рівняння:

$$E[X(t)] = E[X(0)] = \text{Constant} \quad (2.4)$$

б) Чи є дисперсія даного часового ряду константою?

$$\text{Var}[X(t)] = \text{Var}[X(0)] + \sum_{i=1}^t \text{Var}[\varepsilon(i)] \quad (2.5)$$

$$\text{Var}[X(t)] = t * \text{Var}(\text{Error}) = \text{Time dependent} \quad (2.6)$$

Отже, можна зробити висновок, що «випадкове блукання» не є стаціонарним процесом, оскільки воно має дисперсію, яка є функцією від часу. Крім того, якщо б ми ще додатково перевірили коваріацію, то ми б побачили, що вона теж залежить від часу [21].

Таким чином, нам вже відомо, що випадкове блукання – це нестаціонарний процес. Тепер введемо додатковий коефіцієнт  $\rho$  у рівняння (2.1), щоб побачити, чи зможемо ми таким чином перетворити його на рівняння стаціонарного процесу:

$$X(t) = \rho * X(t - 1) + \varepsilon(t) \quad (2.7)$$

Тепер ми будемо змінювати значення коефіцієнта  $\rho$ , щоб побачити, чи зможемо ми таким чином зробити часовий ряд для випадкового блукання стаціонарним. При цьому ми будемо лише візуально інтерпретувати розсіювання, а не проводити будь-які тести для перевірки стаціонарності часового ряду.

Почнемо з ідеально стаціонарного часового ряду при  $\rho = 0$ . На рисунку 2.5 наведено графічне представлення даного часового ряду.

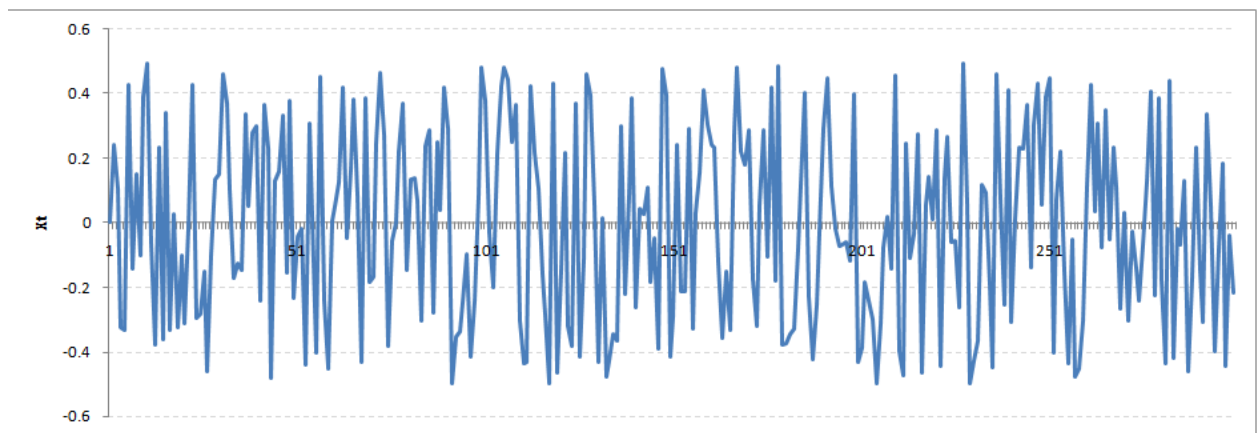


Рисунок 2.5 – Стаціонарний часовий ряд при  $\rho = 0$

Збільшивши значення коефіцієнта  $\rho$  до 0.5, отримаємо часовий ряд із графіком, що зображений на рисунку 2.6.

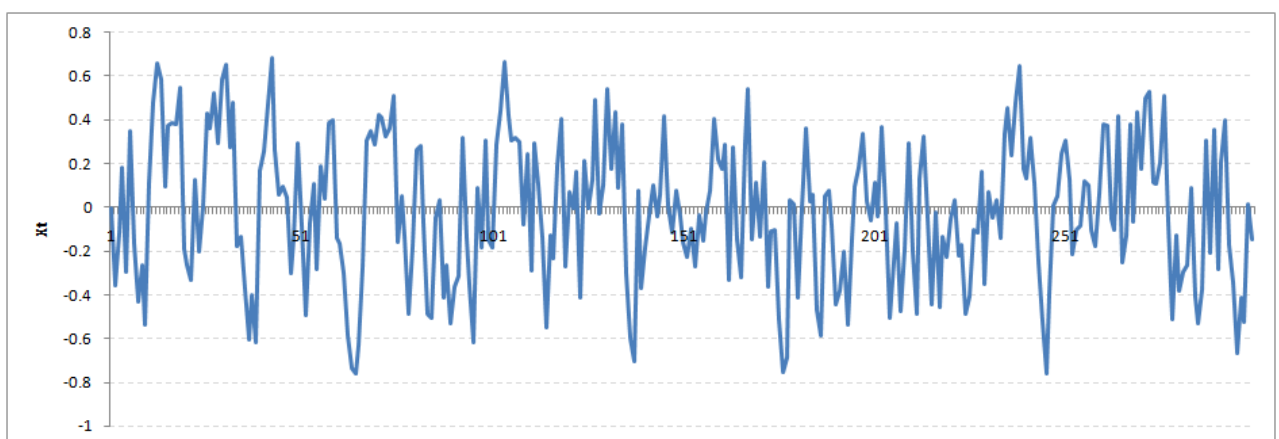


Рисунок 2.6 – Стаціонарний часовий ряд при  $\rho = 0.5$

Як видно із рисунка 2.6, цикли в нашому часовому ряду стали більш протяжними, але по суті на даному графіку не спостерігається серйозних порушень стаціонарності часового ряду.

Тепер розглянемо більш екстремальний випадок, коли  $\rho = 0.9$ . Графічне представлення даного часового ряду зображене на рисунку 2.7.

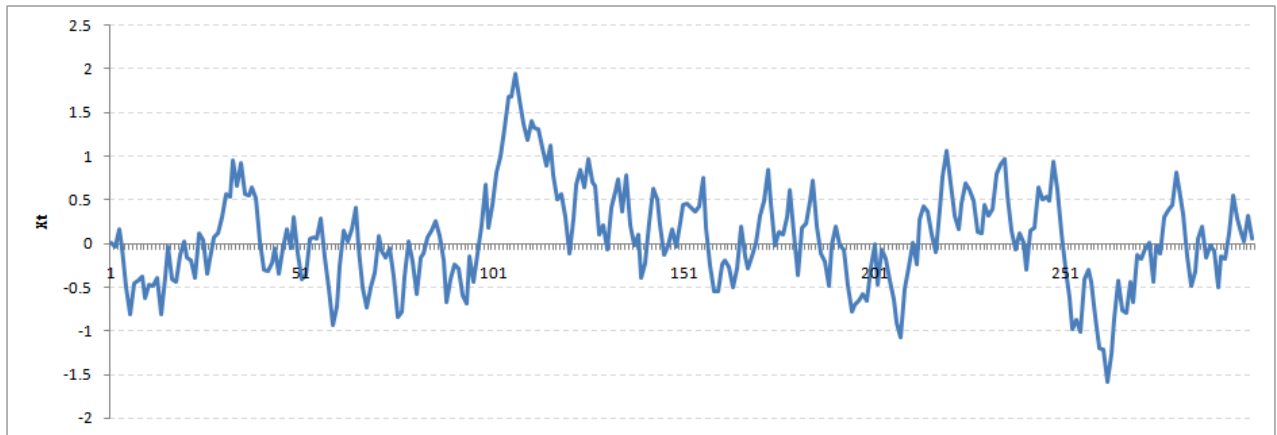


Рисунок 2.7 – Стаціонарний часовий ряд при  $\rho = 0.9$

Як і раніше, ми все ще можемо спостерігати, що  $X$  завжди повертається з екстремальних значень до нуля через певний проміжок часу. Саме тому, даний часовий ряд також не викликає суперечностей із поняттям стаціонарності.

Насамкінець давайте розглянемо «випадкове блукання» при  $\rho = 1$ . Графік даного процесу можна побачити на рисунку 2.8.

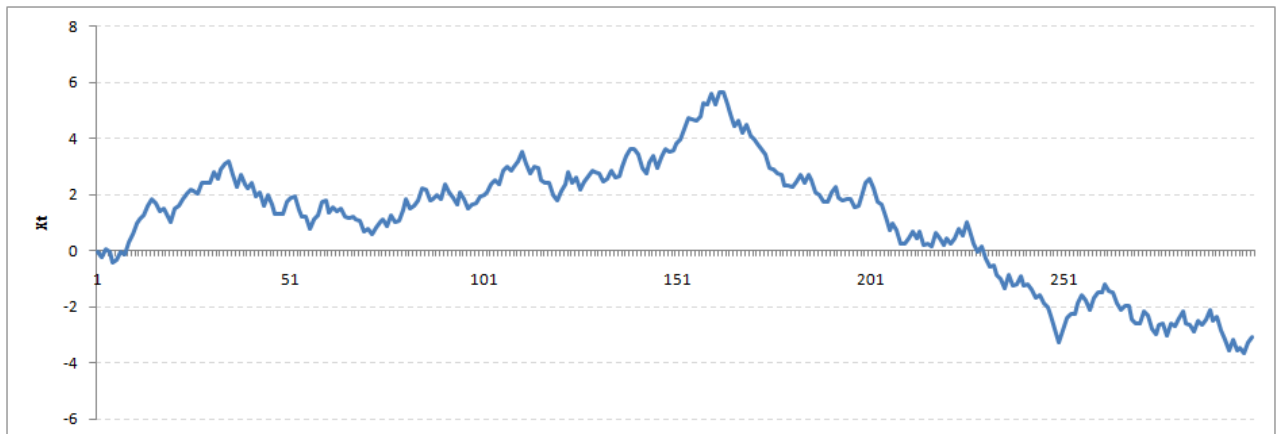


Рисунок 2.8 – Нестационарний часовий ряд при  $\rho = 1$

Очевидно, що в часовому ряді, графік якого зображений на рисунку 2.8, є порушення умов стаціонарності. Тут виникає запитання, а що саме робить випадок  $\rho = 1$  особливим, і при цьому часовий ряд не є стаціонарним? Ми опишемо цю причину із математичної точки зору.

Для того, щоб розібратись у цьому питанні, зробимо наступне. Давайте знайдемо математичне сподівання від кожної сторони рівняння (2.7):

$$E[X(t)] = \rho * E[X(t-1)] \quad (2.8)$$

Це рівняння є дуже інтуїтивним. Наступне значення  $X$  (тобто в момент часу  $t$ ) так би мовити «зменшується» до виразу " $\rho * \text{Останнє значення } X$ ".

Розглянемо наступний приклад:

$$(X(t-1) = 1, \rho = 0.5) \Rightarrow E[X(t)] = 0.5 \quad (2.9)$$

Із формули (2.9) бачимо, що якщо  $X$  рухається в будь-якому напрямку від нуля, він повертають до нуля назад на наступному кроці. Єдиним компонентом, який може керувати ним додатково, є похибка. Як відомо, компонент похибки з однаковою ймовірністю може змінюватись в будь-якому напрямку. То що ж саме відбувається, коли коефіцієнт  $\rho$  стає рівним 1? Ніякий компонент не може «змінити»  $X$  в бік зменшення на наступному кроці. Саме через це стаціонарність втрачається.

Тепер прийшов час описати тести для перевірки часового ряду на стаціонарність. Одним із найпопулярніших таких тестів є тест Дікі-Фуллера. Варто зазначити, що матеріал щодо випадкового блукання та коефіцієнту  $\rho$ , який був розглянутий вище, формально називається тестом Дікі-Фуллера [22].

Для того, щоб привести рівняння (2.7) до вигляду, який використовується в тесті Дікі-Фуллера, потрібно виконати незначне перетворення:

$$\begin{aligned} X(t) &= \rho * X(t-1) + \varepsilon(t) \Rightarrow \\ \Rightarrow X(t) - X(t-1) &= (\rho - 1)X(t-1) + \varepsilon(t) \end{aligned} \quad (2.10)$$

У даному випадку нам потрібно перевірити, вираз  $(\rho - 1)$  істотно відрізняється від нуля, чи ні. Тобто нульова гіпотеза буде мати наступний вигляд:

$$H_0: \rho - 1 = 0 \quad (2.11)$$

Якщо нульова гіпотеза щодо рівності нулю виразу  $(\rho - 1)$  буде відхилена, то ми отримаємо стаціонарний часовий ряд.

Тест на стаціонарність та приведення часового ряду до стаціонарного виду є найбільш критичними процесами під час моделювання часових рядів.

Після приведення часового ряду до стаціонарної форми, можна переходити до процесу моделювання. Найбільш поширеними моделями, що використовуються для моделювання часових рядів, є моделі ARMA (Autoregressive Moving Average Model). У ARMA-моделі AR означає авторегресію, а MA – це ковзне середнє.

Далі будуть описані зазначені вище терміни, та розібрані характеристики, що пов'язані з цими моделями. Але перш за все, слід пам'ятати, що моделі AR та MA не застосовуються для нестаціонарних часових рядів.

Якщо ви маєте справу із нестаціонарним часовим рядом, то вам спочатку потрібно буде перетворити ваш часовий ряд у стаціонарну форму (застосувати метод різниць, видалити тренд або інше підходяще перетворення), а потім, для здійснення моделювання, вибрати з доступних моделей часових рядів найбільш підходящу для конкретного випадку.

В наступних підрозділах буде описана кожна з цих двох моделей (AR та MA) окремо. А також ми розглянемо характеристики цих моделей.

### 2.3 Вибір моделей для прогнозування характеристик трафіка цифрової реклами

Як вже було зазначено раніше, ми маємо справу із трафіком цифрової реклами, що представлений статистичним набором даних. Ці статистичні значення того чи іншого рекламного оголошення фіксувались за однакові інтервали часу. Тобто, ми маємо справу із статистичними даними, що представлені у формі часових рядів. Відповідно до цього ми і будемо обирати моделі для здійснення прогнозування.



У наявному наборі статистичних даних присутня велика кількість характеристик рекламного трафіка. Серед них нам потрібно вибрати декілька ключових характеристик, з якими ми будемо працювати далі. Такими найважливішими характеристиками в нашому контексті є показники доходу та заповнюваності. І саме ці метрики ми будемо прогнозувати.

Оскільки ми маємо справу із часовими рядами, то виберемо моделі часових рядів для здійснення прогнозування. Найпоширенішими прогностичними моделями в області часових рядів є такі моделі: модель авторегресії, модель ковзного середнього, модель авторегресії з ковзним середнім, модель авторегресії з інтегрованим ковзним середнім і модель авторегресії та інтегрованого ковзного середнього з екзогенними змінними.

Для побудови найкращої моделі доцільно скористатись наступною методикою. Потрібно будувати моделі від простішої до складнішої, поступово додаючи до неї певні компоненти, що можуть підвищити її точність. Тобто ми почнемо процес прогнозування із моделі авторегресії, та поступово будемо її ускладнювати.

### 2.3.1 Модель AR

Почнемо розгляд моделі *AR* із прикладів. Для того, щоб краще зрозуміти моделі *AR*, використаємо наступний приклад. Нехай поточне значення ВВП країни, яке позначимо через  $x(t)$ , залежить від ВВП минулого року, тобто  $x(t - 1)$ . Гіпотеза полягає в тому, що загальна вартість виробництва товарів та послуг у країні в бюджетному році (відомий як ВВП) залежить від створених виробничих підприємств / наданих послуг в минулому році та нових товарів / підприємств / послуг в поточному році. Але основна складова ВВП – це все-таки значення за попередній рік.

Отже, ми можемо формально записати рівняння ВВП наступним чином:

$$x(t) = \alpha * x(t - 1) + \varepsilon(t) \quad (2.12)$$

Це рівняння відоме під формулюванням  $AR(1)$ . Цифра (1) означає, що наступне значення виключно залежить від попереднього значення.  $\alpha$  – це коефіцієнт, який ми прагнемо знайти для того, щоб мінімізувати функцію помилки. Зверніть увагу, що  $x(t-1)$  дійсно пов'язаний з  $x(t-2)$  таким же чином. Отже, будь-який стрибок  $x(t)$  поступово згасатиме у майбутньому.

Розглянемо інший приклад. Нехай  $x(t)$  – це кількість пляшок соку, проданих у деякому місті в певний день. Під час зими дуже мало магазинів придбали пляшки з соком. Раптово, у певний день, температура зросла, а попит на пляшки з соком піднявся до 1000 штук. Однак через кілька днів погода знову стала холодною. Але, знаючи, що люди звикли пити соки в теплі дні, у холодні дні 50% людей все ще пили соки. У наступні дні ця частка знизилася до 25% (50% від 50%), а потім поступово впала декількох пляшок в день, після значної кількості днів. Графік, який представлений на рисунку 2.9, пояснює інерційну властивість моделі  $AR$ .

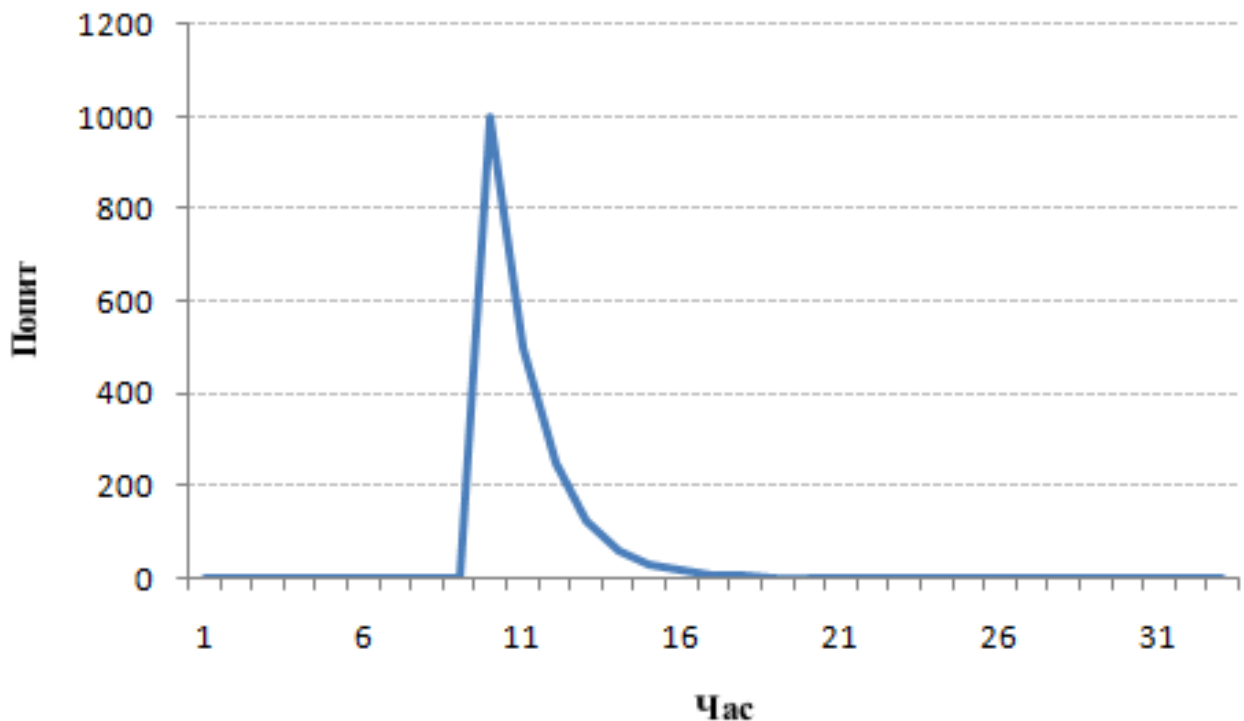


Рисунок 2.9 – Графічне представлення моделі  $AR(1)$

Тепер наведемо загальне рівняння для моделі  $AR(p)$ :

$$y(k) = a_0 + \sum_{i=0}^p a_i y(k-i) + \varepsilon(k), \quad (2.13)$$

де  $\{\varepsilon(k)\}$  – послідовність білого шуму з нульовим середнім,  $E[\varepsilon(k)] = 0$ , та скінченною постійною дисперсією  $E[\varepsilon^2(k)] = \sigma_\varepsilon^2$ ;  $y(0) = y_0$  – детермінована початкова умова.

### 2.3.2 Модель ARMA

Перш ніж перейти до опису моделі *ARMA*, спочатку потрібно розібрати модель *MA*. Тепер давайте розглянемо інший приклад, щоб краще зрозуміти модель ковзного середнього.

Нехай деякий виробник виробляє певний тип сумок, який вже був на ринку. Через велику конкуренцію на ринку, продажі сумок протягом багатьох днів залишались майже нульовими. Отже, одного дня він зробив певний експеримент з дизайном, і виготовив інший тип сумок. Тоді цей тип сумок не був доступний усім на ринку. Таким чином, він зміг продати цілу партію з 1000 сумок (назвемо це як  $x(t)$ ). Попит був настільки високий, що партія закінчилась. Як результат, близько 100 додаткових клієнтів не змогли придбати цю сумку. Давайте називати цю різницю як похибку в даний момент часу. З часом даний тип сумок втратив ефект новинки. Але залишилися ще деякі клієнти, які не змогли придбати дану сумку в попередній день. Таким чином, у формулі (2.14) наведено просте формулювання для відображення даного сценарію:

$$x(t) = \beta * \varepsilon(t-1) + \varepsilon(t) \quad (2.14)$$

Якщо ми спробуємо побудувати графік даного процесу, то він буде виглядати приблизно так, як це зображено на рисунку 2.10.

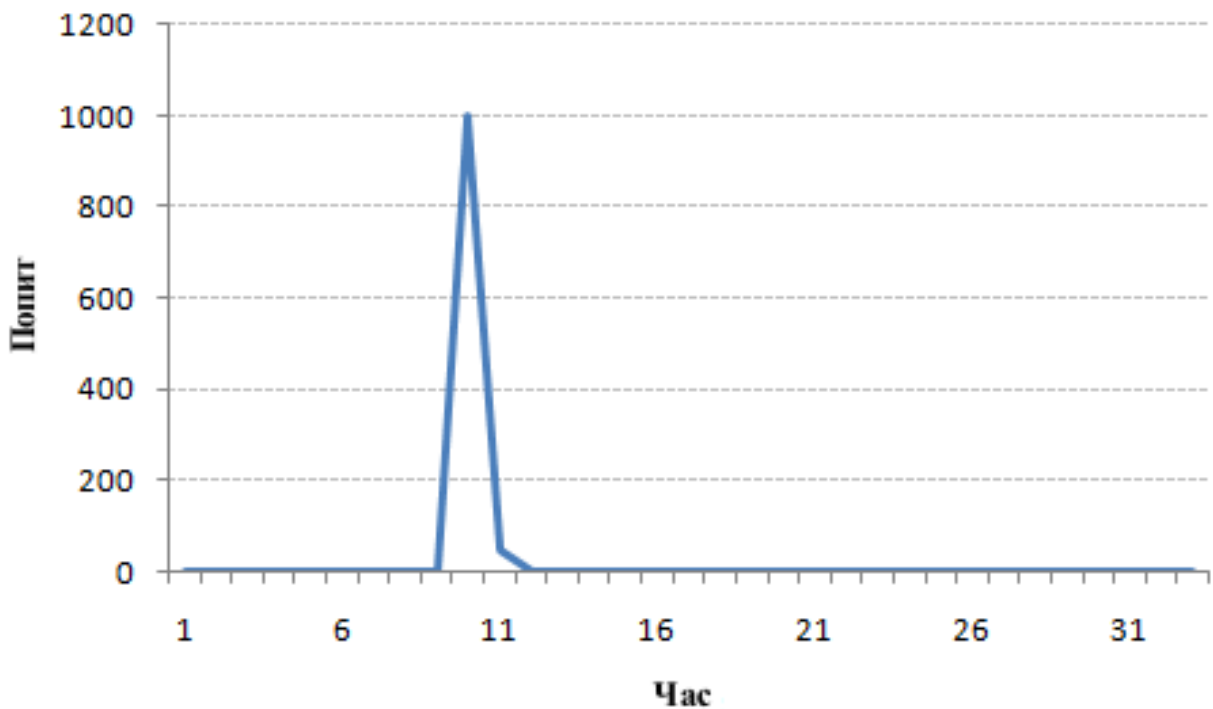


Рисунок 2.10 – Графічне представлення моделі  $MA(1)$

Порівнявши рисунки 2.9 та 2.10, можна помітити різницю між моделлю  $MA$  та  $AR$ . У моделі  $MA$  так звані «стрибки» швидко зникають з часом, при цьому в моделі  $AR$  спостерігається плавне спадання після «стрибка» [23].

Загалом кажучи, основна відмінність моделей  $AR$  та  $MA$  базується на кореляції між елементами часового ряду в різні моменти часу. Кореляція між  $x(t)$  і  $x(t - n)$  при  $n > \text{порядок } MA$  завжди дорівнює нулю. Це безпосередньо випливає з того, що для моделей  $MA$  коваріація між  $x(t)$  і  $x(t - n)$  дорівнює нулю (це, зокрема, пояснюється прикладом, який був наведений на рисунку 2.9). Проте для моделей  $AR$  кореляція між  $x(t)$  і  $x(t - n)$  поступово зменшується із збільшенням  $n$ . Ця різниця використовується незалежно від того, яка модель у вас є –  $AR$  чи  $MA$ . Також варто зауважити, що із графіку кореляції ми можемо отримати порядок моделі  $MA$ .

Отже, тепер потрібно дещо узагальнити підходи, що описані вище. Коли ми отримали стаціонарний часовий ряд, то ми повинні відповісти на два основних запитання:

- а) це процес  $AR$  чи  $MA$ ?

б) який порядок процесу  $AR$  або  $MA$  нам потрібно використовувати?

На перше запитання можна відповісти, використовуючи загальну кореляційну діаграму (також відома як Auto Correlation Function /  $ACF$ ).  $ACF$  – це графік загальної кореляції між різними функціями затримки. Наприклад, у наведеному вище прикладі з  $ВВП$ ,  $ВВП$  у момент часу  $t$  – це  $x(t)$ . Ми зацікавлені в тому, щоб знайти кореляцію  $x(t)$  з  $x(t - 1)$ ,  $x(t - 2)$  і так далі. Тепер потрібно врахувати ті твердження, що наводились вище.

У моделі ковзного середнього з порядком  $n$  ми не отримаємо ніякої кореляції між  $x(t)$  і  $x(t - n - 1)$ . Отже, загальна кореляційна діаграма в даному випадку різко «падає» після запізнення порядку  $n$ . Тому стає досить просто знайти порядок для моделі  $MA$ . Для моделі  $AR$  ця кореляція поступово зменшуватиметься без стрибка. То що ж тоді робити, якщо це маємо справу з моделлю  $AR$ ?

В такому випадку на допомогу приходить інший підхід. Якщо знайти часткову кореляцію кожного відставання, то вона буде різко «падати» після порядку моделі  $AR$ . Наприклад, якщо у нас є модель  $AR(1)$ , і при цьому ми виключимо ефект 1-го лагу ( $x(t - 1)$ ), наш другий лаг ( $x(t - 2)$ ) вже не буде залежати від  $x(t)$ . Отже, часткова кореляційна функція ( $PACF$ ) різко падає після першого лагу. Нижче, на рисунках 2.11, 2.12, наведені приклади, які прояснюють будь-які сумніви щодо цієї концепції.

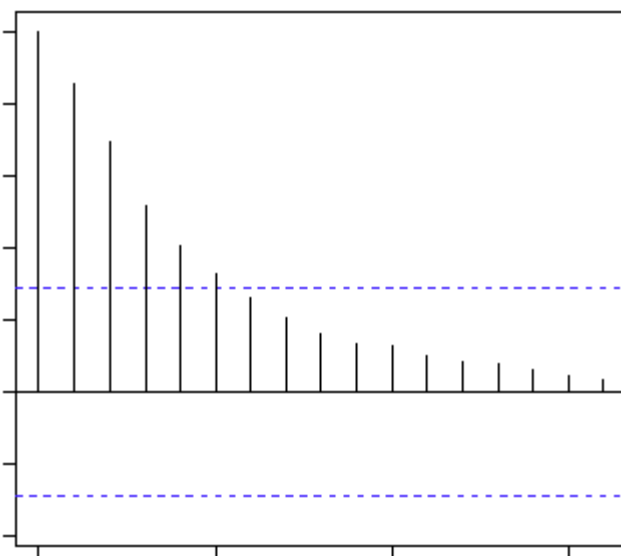


Рисунок 2.11 –  $ACF$  для  $AR(2)$

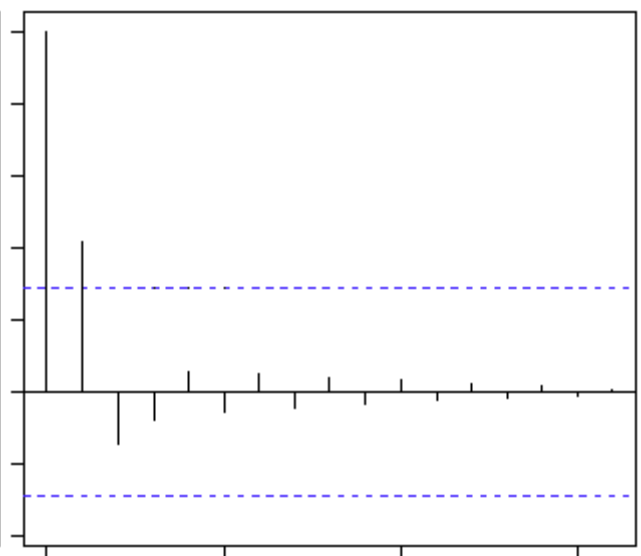


Рисунок 2.12 –  $PACF$  для  $AR(2)$

Синя лінія на рисунках 2.11, 2.12 показує значення, які суттєво відрізняються від нуля. Якщо ми подивимось на рисунок 2.12, то побачимо що PACF різко падає після 2-го лага, що означає, що це, швидше за все, модель  $AR(2)$ .

Тепер на рисунках 2.13, 2.14 наведемо інший приклад, на цей раз вже із моделлю  $MA$ .

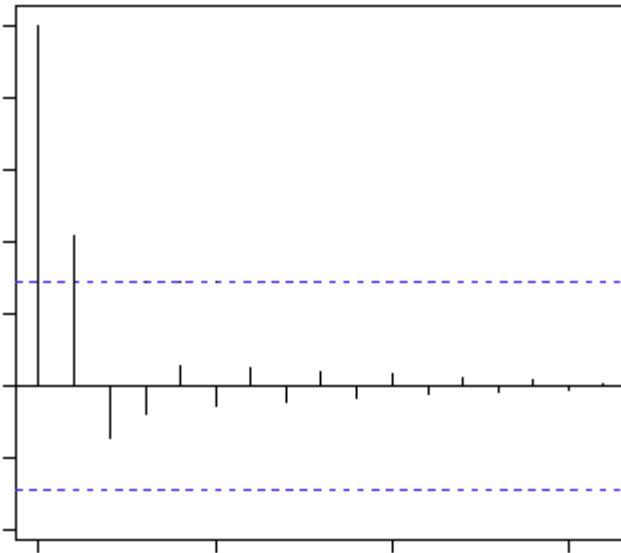


Рисунок 2.13 –  $ACF$  для  $MA(2)$

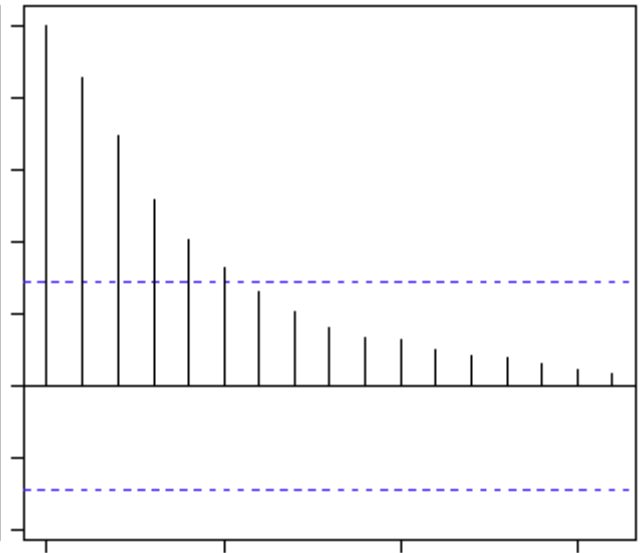


Рисунок 2.14 –  $PACF$  для  $MA(2)$

Як бачимо із рисунка 2.13, графік  $ACF$  різко падає після 2-го лага, що означає, що це, швидше за все, модель  $MA(2)$ .

Таким чином, ми розглянули, як визначити тип стаціонарного часового ряду, використовуючи графіки  $ACF$  та  $PACF$ .

Тепер наведемо загальне рівняння для моделі  $ARMA(p, q)$ :

$$y(k) = a_0 + \sum_{i=1}^p a_i y(k-i) + \sum_{j=1}^q b_j v(k-j) + \varepsilon(k) \quad (2.15)$$

де  $p, q$  – порядки авторегресії та ковзного середнього відповідно,  $\varepsilon(k)$  – випадкова величина, поява якої зумовлена впливом випадкових збурень на процес, який моделюється.

### 2.3.3 Модель ARIMA

Модель *ARIMA* – це досить поширений клас статистичних моделей для аналізу та прогнозування даних часових рядів. Вона добре працює із стандартними структурами в даних у формі часових рядів, і таким чином забезпечує простий, але потужний спосіб для якісних прогнозів часових рядів.

*ARIMA* – це аббревіатура, що означає AutoRegressive Integrated Moving Average. Це узагальнення простішої моделі AutoRegressive Moving Average, що додає поняття інтегрованості.

Дана аббревіатура є описовою, що фіксує ключові аспекти самої моделі. До моделі *ARIMA* входять наступні складові:

- *AR*: авторегресія. Модель, яка використовує зв'язок між деяким спостереженням та певною кількістю спостережень із затримкою;
- *I*: інтегрованість. Застосування методу різниць до спостережень (наприклад, віднімання спостереження від спостереження на попередньому кроці), щоб зробити часовий ряд стаціонарним;
- *MA*: ковзне середнє. Модель, яка використовує залежність між спостереженням та залишковою помилкою від моделі ковзного середнього, застосованої до спостережень із лагами.

Кожен із цих компонентів явно вказаний у моделі як параметр. Стандартне позначення моделі –  $ARIMA(p, d, q)$ , де параметри замінюються цілими значеннями, щоб швидко вказати конкретну модель *ARIMA*, яка використовується.

Параметри моделі *ARIMA* визначаються наступним чином:

- *p*: кількість спостережень із затримкою, включених у модель, також називається порядком затримки;
- *d*: кількість разів, коли до спостережень застосовують метод різниць, також називають порядком різниць;
- *q*: розмір вікна ковзного середнього, також називається порядком ковзного середнього.

Далі будується лінійна регресійна модель, враховуючи вказані вище параметри, а до часового ряду також застосовують метод різниць, щоб зробити його стаціонарним, тобто видалити тренди та сезонні компоненти, які негативно впливають на регресійну модель.

Значення 0 також може використовуватися для певного параметра, причому це вказує на те, що конкретний компонент не слід використовувати в даній моделі. Таким чином, модель *ARIMA* може бути налаштована на виконання функцій моделі *ARMA*, і навіть простіших моделей *AR*, *I* або *MA*.

Застосування моделі *ARIMA* до часового ряду передбачає, що основним процесом, який генерує спостереження, є процес *ARIMA*. Це може здатися очевидним, але допомагає обґрунтовувати необхідність підтвердження припущень моделі для початкових спостережень та залишкових помилок прогнозів моделі.

Тепер пояснимо в чому саме полягає відмінність моделі *ARIMA* від моделі *ARMA*. Розглянемо спочатку рівняння для моделі *ARMA*.

Рівняння виду:

$$\begin{aligned} y(k) &= a_0 + \sum_{i=1}^p a_i y(k-i) + \sum_{j=1}^q b_j \varepsilon(k-j) + \varepsilon(k) = \\ &= a_0 + \sum_{i=1}^p a_i y(k-i) + \sum_{j=0}^q b_j \varepsilon(k-j), \end{aligned} \quad (2.16)$$

де  $b_0 = 1$ ,  $\sum_{j=1}^q b_j = 1$ ,

називають моделлю авторегресії та ковзного середнього (*ARMA*), якщо корені його характеристичного рівняння лежать всередині одиничного кола на комплексній площині, а  $\{\varepsilon(k)\}$  – випадковий процес.

При цьому, якщо один або більше коренів характеристичного рівняння, записаного для (2.16), дорівнюють одиниці, то послідовність  $\{y(k)\}$  називають



інтегрованою або процесом з одиничними коренями, а рівняння (2.16) називають авторегресією з інтегрованим ковзним середнім (*ARIMA*). Процеси з одиничними коренями відносяться до класу нестационарних процесів. Оскільки такі процеси є досить характерними для виробничих технологій, економіки, фінансів, екології та інших галузей, то їм необхідно приділяти значну увагу.

#### 2.3.4 Модель *ARIMAX*

Стандартна модель *ARIMA* (autoregressive integrated moving average) дозволяє робити прогнози, спираючись лише на минулі значення прогнозованої змінної. Модель передбачає, що майбутні значення змінної лінійно залежать від її минулих значень, а також від значень минулих (стохастичних) похибок. Модель *ARIMAX* – це розширена версія моделі *ARIMA*. Вона додатково включає в себе незалежні змінні (регресори). Дана модель також називається як динамічна регресійна модель [24].

Модель *ARIMAX* схожа на багатовимірну регресійну модель, але дозволяє скористатися перевагами автокореляції, яка може бути присутня в залишках регресії, для підвищення точності прогнозу.

Отож, давайте тепер розглянемо модель авторегресії та інтегрованого ковзного середнього з екзогенними змінами (*ARIMAX*). Не зважаючи на складну та заплутану назву, дана модель є досить популярною і часто використовується для аналізу часових рядів. По суті, це просто комбінація з 3 моделей: авторегресійної моделі, що використовує попередні стани; моделі ковзного середнього, що використовує попередні залишки; та звичайної регресійної моделі, що використовує зовнішні змінні, на інтегрованому часовому ряді. Зазвичай ця модель записується як  $ARIMAX(p, d, q, b)$ .

Для того, щоб краще зрозуміти дану модель, наведемо для неї узагальнену формулу:

$$y(k) = a_0 + \sum_{i=1}^p a_i y(k-i) + \sum_{j=1}^q b_j v(k-j) + \sum_{h=1}^b c_h x_h + \varepsilon(k), \quad (2.17)$$

де  $p, q$  – порядки авторегресії, ковзного середнього відповідно.  $b$  – кількість включених пояснюючих змінних.

Варто зазначити, що в даній моделі  $d$  – це порядок тренду, тобто інтегрованої складової. В рівнянні (2.17) сума  $\sum_{h=1}^b c_h x_h$  являє собою лінійну комбінацію зовнішніх пояснюючих змінних.

Рішення про включення в модель *ARIMAX* відповідної пояснюючої змінної  $x_h$  приймається на основі аналізу сумісної кореляції вихідної змінної  $y$  та  $x_h$ . Якщо ця кореляція більша за певне порогове значення, то змінну  $x_h$  необхідно включати до моделі.

#### 2.4 Критеріальна база для аналізу якості результатів та вибору кращої моделі

Після того, як обрані моделі будуть побудовані, виникає необхідність порівняти їх між собою. Це потрібно зробити для того, щоб вибрати найкращу модель, та використовувати саме її для подальших досліджень.

Тому наведемо перелік критерії та показників, за якими буде вибиратись найкраща модель [25]:

- статистика Дарбіна-Уотсона ( $DW$ ). Статистика Дарбіна-Уотсона ( $DW$  – Durbin-Watson) показує адекватність побудованої моделі та обчислюється за формулою (2.18):

$$DW = \frac{\sum_{t=2}^N (\varepsilon(t) - \varepsilon(t-1))^2}{\sum_{t=1}^N \varepsilon^2(t)}, \quad (2.18)$$

де  $\varepsilon$  – вектор залишків (різниця між значеннями отриманими за моделлю та фактичними), при цьому  $DW \in [0; 4]$ . Для найкращої моделі  $DW \rightarrow 2$ , це означає що залишки моделі між собою не корелюють;

- коефіцієнт множинної детермінації  $R^2$ . Даний коефіцієнт розраховується за формулою (2.19):

$$R^2 = \frac{\text{var}(\hat{y})}{\text{var}(y)} = 1 - \frac{SSE}{SST} \quad (2.19)$$

де  $\text{var}(\hat{y})$  – дисперсія залежної змінної, що оцінена із використанням побудованої моделі;  $\text{var}(y)$  – дисперсія вимірів залежної змінної;  $SSE = \sum_{k=1}^N [y(k) - \hat{y}(k)]^2$  – сума квадратів похибок (залишків) моделі (sum of squared errors);  $SST = \sum_{k=1}^N [y(k) - \bar{y}]^2$  – загальна сума квадратів (total sum of squares);  $\bar{y}$  – середнє значення;  $SST = SSE + SSR$ , де  $SSR = \sum_{k=1}^N [\hat{y}(k) - \bar{y}]^2$  – загальна сума квадратів для регресії (sum of squares for regression).

Очевидно, що найкращим значенням є  $R^2 = 1$ , тобто, коли дисперсії вимірів змінної, та цієї ж змінної, оціненої за рівнянням, збігаються. Цей коефіцієнт можна трактувати, також, як міру інформативності моделі, якщо за міру інформативності вибрати дисперсію. Таким чином,  $R^2$  показує рівень інформативності моделі по відношенню до інформативності вибірки даних, за допомогою якої вона була оцінена;

- критерій Акайке ( $AIC$ ). Інформаційний критерій Акайке використовують для оцінки адекватності моделі. Його можна розрахувати за формулою (2.20):

$$AIC = N \ln \left( \sum_{k=1}^N e^2(k) \right) + 2n, \quad (2.20)$$

де  $n = p + q + 1$  – кількість параметрів моделі, що оцінюються за допомогою статистичних даних ( $p$  – число параметрів авторегресійної частини моделі;  $q$  – число параметрів ковзного середнього; 1 з'являється тоді, коли оцінюється зміщення (або перетин, тобто  $a_0$ )).

Критерій Акайке містить в правій частині суму квадратів похибок, а тому за цим критеріями вибирають ту модель, для якої критерій приймає найменші значення. Введення нового регресора приводить до збільшення критерію (при цьому збільшується  $n$ ), але одночасно зменшується сума квадратів похибок і критерій в цілому зменшується. Якщо регресор не покращує модель, то критерій збільшується;

- байєсівський інформаційний критерій ( $BIC$ ). Інформаційний критерій Байєса, так само як і критерій Акайке, використовують для оцінки адекватності побудованих моделей. Даний критерій розраховується за формулою (2.21):

$$BIC = N \ln \left( \sum_{k=1}^N e^2(k) \right) + n \ln(N), \quad (2.20)$$

де параметри  $n, p, q$  – аналогічні до тих що були в критерії Акайке.

У інформаційного критерія Байєса досить схожі властивості до тих, що були в критерія Акайке. Тобто для кращої моделі  $BIC \rightarrow \min$ . Проте варто зазначити, що асимптотичні властивості для довгих вибірок кращі у критерія Байєса, тобто, його рекомендують застосовувати при відносно великих значеннях  $N$  ( $N > 100$ ).

Також наведемо перелік найбільш популярний похибок, що використовуються при порівнянні моделей [25]:

- *RMSE*. Середньоквадратична похибка або корінь із квадрата середньої похибки (в літературі *RMSE* – root mean square error) в загальному випадку обчислюється за формулою (2.21):

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \quad (2.21)$$

де  $N$  – розмір вибірки,  $y$  – фактичні значення, а  $\hat{y}$  – значення побудовані за моделлю;

- *MAE*. Середня абсолютна похибка (MAE) – визначає середню величину помилок у наборі прогнозів, не враховуючи їх напрямки. Це середнє значення, для тестової вибірки, абсолютних різниць між прогнозом та фактичним спостереженням, де всі індивідуальні різниці мають однакову вагу. Дана похибка розраховується за формулою (2.22):

$$MAE = \frac{1}{N} \sum_{j=1}^N |y_j - \hat{y}_j|, \quad (2.22)$$

де  $N$  – розмір вибірки,  $y$  – фактичні значення, а  $\hat{y}$  – значення побудовані за моделлю.

Якщо у формулі не застосовується модуль (знак помилок не видаляються), середня помилка стає середньою зміщеною помилкою (*MBE*) і, як правило, призначена для вимірювання середнього зміщення моделі. *MBE* може передавати корисну інформацію, але слід тлумачити її обережно, оскільки позитивні та негативні помилки можуть взаємовиключатись;

- *MAPE*. Середня абсолютна похибка у відсотках (*MAPE*) – найпоширеніша міра помилки для прогнозів. *MAPE* найкраще працює, коли дані не містять екстремальних значень (включаючи нулі).

З нулями або значеннями, близькими до нуля, *MAPE* може дати спотворене відображення помилки. Помилка на майже нульовому об'єкті може бути нескінченно високою, що призводить до спотворення загальної помилки, коли вона усереднюється. Для прогнозів об'єктів, що знаходяться поблизу нуля, симетрична середня абсолютна відсоткова помилка (*SMAPE*) є кращим рішенням.

*MAPE* – це середня абсолютна помилка у відсотках для кожного періоду часу або прогнозу, за вирахуванням фактичних даних, поділена на фактичні дані. Розрахунку даної похибки наведений у формулі (2.23):

$$MAPE = \frac{1}{N} \sum_{j=1}^N \frac{|y_j - \hat{y}_j|}{|y_j|} \cdot 100\%, \quad (2.22)$$

де  $N$  – розмір вибірки,  $y$  – фактичні значення, а  $\hat{y}$  – значення побудовані за моделлю;

- *MASE*. Середня абсолютна масштабована похибка (*MASE*) – це показник точності прогнозів. Дана похибка має деякі сприятливі властивості в порівнянні з іншими методами для обчислення помилок прогнозування, такими як середньоквадратичне відхилення, і тому рекомендується для визначення порівняльної точності прогнозів. Серед переваг даної похибки можна виділити наступні властивості: масштабна інваріантність; передбачувана поведінка при  $y_t \rightarrow 0$ ; симетричність; інтерпретовність; асимптотична нормальність [26].

Для розрахунку похибки *MASE*, потрібно використати допоміжний вираз, що наведений у формулі (2.23):

$$Q = \frac{1}{N-1} \sum_{j=2}^N |y_j - y_{j-1}|, \quad (2.23)$$

де  $N$  – розмір вибірки,  $y$  – фактичні значення.

Варто зазначити, що  $Q$  – це статистика для масштабування, що обчислена за тренувальними даними. Для не сезонних часових рядів, зручний спосіб визначення статистики масштабування – це середня абсолютна похибка між послідовними спостереженнями. Тобто  $Q$  – це  $MAE$  для найвних прогнозів, що обчислюються за тренувальними даними. Оскільки похибки  $Q$  та  $MAE$  обидві включають значення в масштабі вихідних даних, то їхнє співвідношення не залежить від шкали даних. Масштабована помилка менша від одиниці, якщо кращим є прогноз, побудований за моделлю, а не середнє значення прогнозу, розрахованого за навчальними даними. І навпаки, вона більша, ніж одиниця, якщо модельний прогноз гірший, ніж середнє значення найвного прогнозу, який обчислюється за тренувальними даними.

Тепер наведемо власне формулу, за якою розраховується похибка  $MASE$ . Розрахунки досить прості та наведені у формулі (2.24):

$$MASE = \frac{MAE}{Q}, \quad (2.24)$$

де  $MAE$  – середня абсолютна похибка,  $Q$  – статистика для масштабування, що розраховується за формулою (2.23).

Значення статистики  $Q$  обчислюється з використанням тренувальних даних, оскільки важливо отримати стабільну міру для масштабу даних. Набір тренувальних даних зазвичай набагато більший, ніж тестовий набір, і саме тому дозволяє краще оцінити  $Q$ .

## Висновки до розділу 2

Отже, в даному розділі магістерської дисертації було описано методику моделювання та прогнозування характеристик трафіка цифрової реклами. Зокрема, було виконано наступні процедури: проаналізовано та конкретизовано поставлені задачі; описано методику побудови математичних моделей на основі статистичних даних у формі часових рядів; проведений підбір моделей для прогнозування характеристик рекламного трафіка; описано критеріальну базу для вибору кращої моделі.

Також були наведені основні поняття теорії моделювання та прогнозування часових рядів. А саме – були описані такі поняття: часовий ряд, випадкове блукання, стаціонарність часового ряду, тест Дікі-Фуллера, модель авторегресії (*AR*), модель ковзного середнього (*MA*), автокореляційна функція (*ACF*), часткова автокореляційна функція (*PACF*).

Далі були описані більш складні моделі часових рядів. Такі як модель авторегресії та ковзного середнього (*ARMA*), модель авторегресії та інтегрованого ковзного середнього (*ARIMA*), а також модель авторегресії та інтегрованого ковзного середнього з екзогенними змінними (*ARIMAX*).

Насамкінець, було описано критеріальну базу для аналізу якості результатів та вибору кращої моделі. Зокрема, було наведено найпоширеніші критерії та показники для порівняння моделей, а саме: статистика Дарбіна-Уотсона (*DW*), коефіцієнт множинної детермінації  $R^2$ , інформаційний критерій Акайке (*AIC*), байєсівський інформаційний критерій (*BIC*). Також було описано найбільш популярні похибки, що використовуються при аналізі точності моделей часових рядів. А саме, були наведені наступні похибки: середньоквадратична похибка (*RMSE*), середня абсолютна похибка (*MAE*), середня абсолютна похибка у відсотках (*MAPE*), середня абсолютна масштабована похибка (*MASE*).

Зазначені критерії та похибки дають змогу обрати кращу модель із побудованих. Саме цю модель потрібно використовувати на практиці.



## РОЗДІЛ 3 ПОБУДОВА МОДЕЛЕЙ ПРОГНОЗУ ВАЖЛИВИХ ХАРАКТЕРИСТИК ТРАФІКА ЦИФРОВОЇ РЕКЛАМИ

### 3.1 Вибір даних для обробки та формулювання конкретних задач

#### 3.1.1 Опис статистичний даних для прогнозування

Набір статистичних даних для прогнозування був взятий із статистики рекламного трафіку певної рекламної компанії-посередника між рекламодавцями та видавцями.

Опишемо ще раз постановку задачі. Деяка рекламна компанія-посередник щоденно збирає статистичні дані щодо рекламного трафіку який вона обробляє. У цьому наборі даних міститься багато параметрів, які певним чином впливають на декілька основних, а саме – дохід та показник заповнюваності («fill rate» – співвідношення кількості показів реклами до загальної кількості запитів). Задача полягає у побудові моделей прогнозу доходу та заповнюваності на основі наявних параметрів.

Тепер опишемо вихідні дані, що були обрані для виконання прогнозування. Набір даних містить статистичні показники для рекламного трафіку за 7 місяців – з вересня 2017 року по березень 2018 року. У даному наборі містяться спостереження за 212 днів, причому цей набір має 2 інформаційні параметри – дата та назва рекламного тегу, а також 17 числових статистичних показників.

Наведемо детальний опис атрибутів:

– інформаційні атрибути:

- 1) 1 – Дата (текстовий тип даних);
- 2) 2 – Назва рекламного тегу (текстовий тип даних).

– вхідні змінні:

- 1) 3 – Кількість запитів (числовий тип даних);
- 2) 4 – Кількість запитів відповідної реклами (числовий тип даних);
- 3) 6 – Кількість кліків по рекламі (числовий тип даних);
- 4) 7 – Метрика *CTR* для запитів: *Clicks / Ad requests* (числовий тип даних);

- 5) 8 – Метрика *CTR* для зіставлених запитів: *Clicks / Matched requests* (числовий тип даних);
  - 6) 9 – Метрика *CTR* для показів: *Clicks / Ad impressions* (числовий тип даних);
  - 7) 10 – Ціна за клік (числовий тип даних);
  - 8) 12 – Кількість показів (числовий тип даних);
  - 9) 13 – Метрика *CPM*: *Revenue / Ad Impressions \* 1000* (числовий тип даних);
  - 10) 14 – Метрика *Skip Ratio*: *Skips / Ad impressions* (числовий тип даних);
  - 11) 15 – Кількість переглядів реклами, яку можна пропустити (числовий тип даних);
  - 12) 16 – Метрика *VTR*: *video ad skippable views / Ad impressions* (числовий тип даних);
  - 13) 17 – Метрика *active\_view\_eligible\_count* (числовий тип даних);
  - 14) 18 – Метрика *active\_view\_measurable\_count* (числовий тип даних);
  - 15) 19 – Метрика *active\_view\_viewable\_count* (числовий тип даних).
- вихідні змінні:
- 1) 5 – Заповнюваність: *Matched requests / Ad requests* (числовий тип даних);
  - 2) 11 – Дохід (числовий тип даних).

Кількість незалежних змінних залежить від прогнозованого параметру. Так, наприклад, при прогнозуванні доходу матимемо 16 незалежних параметрів та 1 залежний параметр. У випадку із заповнюваністю ситуація дещо інша. По-перше, у даному наборі даних цей показник обчислюється трохи інакше ніж зазвичай. А саме – *Matched requests / Ad requests*. Очевидно, що потрібно виключити наведені параметри при побудові моделей. Таким чином, матимемо 14 незалежних параметрів та 1 залежний.

Перед тим, як переходити до моделювання, потрібно навести опис основних понять та характеристик цифрової реклами. Зокрема, опишемо ті характеристики, що наявні в нашому наборі даних:

- *date* – дата, коли було зроблене спостереження рекламного оголошення;
- *ad tag name* – назва рекламного тегу;
- *ad requests* – кількість запитів реклами;
- *matched ad requests* – кількість зіставлених запитів реклами;
- *clicks* – кількість кліків по рекламі;
- *earnings* – дохід від рекламного оголошення;
- *ad impressions* – кількість показів рекламного оголошення;
- *active view eligible count* – кількість показів, які повинен вимірювати інструмент «Active View»;
- *active view measurable count* – кількість показів, які реально може виміряти інструмент «Active View»;
- *active view viewable count* – кількість показів, які інструмент «Active View» позначив як справжні;
- *ad requests coverage* –  $\text{matched ad requests} / \text{ad requests}$ ;
- *ad requests ctr* –  $\text{clicks} / \text{ad requests}$ ;
- *matched ad requests ctr* –  $\text{clicks} / \text{matched ad requests}$ ;
- *individual ad impressions ctr* –  $\text{clicks} / \text{ad impressions}$ ;
- *cost per click* – ціна за клік;
- *ad impressions rpm* – ціна за тисячу показів оголошення;
- *video ad skippable views* – кількість переглядів відео-реклами, яку можна пропустити;
- *video ad skippable skip ratio* –  $\text{skips} / \text{ad impressions}$ . Застосовується тільки до відео-реклами, яку можна пропустити;
- *video ad skippable vtr* –  $\text{video ad skippable views} / \text{ad impressions}$ . Застосовується тільки до відео-реклами, яку можна пропустити.

### 3.1.2 Постановка конкретних задач для проведення моделювання

Бізнес-задача, що сформульована вище, зводиться до задачі побудови регресійних моделей прогнозу параметрів доходу та заповнюваності. Це можна пояснити також тим, що статистичні дані збираються компанією через рівні інтервали часу, тобто можна зробити висновок, що ми маємо справу із часовими рядами. А вже для часових рядів потрібно підбирати моделі прогнозування, серед яких одні із найпоширеніших – регресійні.

Згідно із постановкою задачі, у даній роботі потрібно побудувати дві моделі прогнозу для різних статистичних параметрів цифрового рекламного трафіку. Зокрема, перша модель – це модель прогнозу доходу, а друга модель – це модель прогнозу параметра заповнюваності. У даному під заповнюваністю будемо розуміти метрику, що наведена у формулі (3.1):

$$\frac{\text{кількість зіставлених запитів}}{\text{загальна кількість запитів}} \cdot 100\% . \quad (3.1)$$

Тобто маємо справу із відсотком зіставлених запитів у загальній кількості запитів що були надіслані. У свою чергу зіставлений запит – це такий запит реклами, для якого відповідна реклама знайшлась, тобто запиту можна зіставити певну рекламу.

Перед тим як приступити до моделювання, потрібно виконати ще декілька важливих етапів. А саме, спочатку потрібно відповідним чином підгодувати вхідні дані. Зокрема, потрібно зазначити, що набір даних, який був взятий до аналізу дещо відрізняється від того, який нам потрібен для виконання моделювання. Опишемо проблему більш детально. У початковому наборі даних є 19 параметрів, із яких 2 інформаційні параметри та 17 числових параметрів, що являють собою статистичні метрики для трафіка цифрової реклами. При цьому, варто зазначити, що для виконання прогнозування потрібний агрегований набір даних по дням. Тобто треба виконати агрегацію вихідного набору даних по дням, при цьому параметри, що являють собою кількість чогось, потрібно агрегувати

за допомогою суми, а ті параметри, що є певним відсотковим відношенням, або, як винятки: ціна за клік та ціна за тисячу показів рекламного оголошення, потрібно агрегувати за допомогою середнього значення. Оскільки початкове розбиття вихідного набору даних було здійснене за параметрами дати та назви рекламного тегу, а в результаті нам потрібно отримати агреговані дані по дням, то відбудеться агрегація статистичного набору даних по рекламних тегах для кожного дня. Тип агрегації був вказаний вище. При цьому, після агрегації інформаційний параметр «назва рекламного тегу» зникне, і у нас залишиться лише один інформаційний параметр – це дата показу рекламного оголошення.

Також не потрібно забувати про такі важливі етапи підготовки статистичних даних до аналізу як очищення даних та приведення типів. Зокрема, спочатку потрібно провести процедуру приведення типів даних статистичних метрик. Вона полягає у тому, щоб для відповідних параметрів були підібрані найбільш підходящі типи даних. Тобто для параметрів, які характеризують кількість чогось, потрібно використати цілочисельний тип даних, а для параметрів що являють собою певне співвідношення або деяку ціну – потрібно використати тип даних для дійсних чисел, тобто із десятковою крапкою. Окремо можна згадати про параметри «дата» та «назва рекламного тегу». Для дати потрібно використати тип даних, що характеризує дату, для назви рекламного тегу – текстовий тип даних. Після того, як описані вище процедури будуть виконані, потрібно приступити до очищення даних. А саме – потрібно обробити порожні значення. У нашому випадку ми будемо замінювати порожні значення на «0», оскільки це буде означати, що даний рекламний тег є неактивним, про що власне кажучи і свідчить порожнє значення для певного статистичного параметру рекламного тегу. Після цього потрібно виконати очищення набору даних від так званих «викидів», тобто ймовірно помилкових значень, що були записані в поле певної метрики. У нашому випадку було виявлено наступні «викиди»: значення статистичних метрик, які менші від нуля. Ці значення є помилковими, оскільки всі метрики статистичного рекламного трафіку цифрової реклами апріорі є невід’ємними. Тому такі значення потрібно виявляти ще до

проведення моделювання, оскільки вони можуть привести до неправильних висновків у процесі моделювання. Таким чином, було вирішено замінювати такі значення також на «0». Це досить логічно, оскільки помилкові значення свідчать про те, що даний тег насправді не використовувався. Тому заміна на значення «0» виглядає досить природньо, вона не змінює значення даних, і якраз свідчить про те, що рекламний тег не є активним, що привело нас до попереднього пункту. Після того, як усі ці підготовчі етапи статистичних даних буде виконано, можна переходити до безпосередньо моделювання.

Але варто зазначити, що у вихідному наборі даних кількість параметрів є досить великою, і на практиці для побудови моделей це не завжди добре. Тому досить логічно на наступному кроці виконати аналіз значущості параметрів. Оскільки у даній роботі виконується побудова двох моделей прогнозу, то потрібно також проводити аналіз значущості параметрів окремо для кожного випадку. Як було зазначено вище, ми здійснили агрегацію вихідного набору даних, і тепер залишилось вже не 19 параметрів, а 18, тому що атрибут «назва рекламного тегу» зник в результаті агрегування даних по дням. Тепер опишемо як буде виконуватись аналіз значущості наявних параметрів. Для моделі прогнозу доходу будемо діяти наступним чином. В агрегованому наборі даних залишилось 17 числових статистичних характеристик та 1 інформаційний атрибут. При цьому, із цих 17 параметрів один параметр – власне сам дохід. Тому будемо в даному випадку буде перевірятись значущість 16 параметрів по відношенню до доходу, оскільки інформаційні атрибути у даній процедурі відкидаються. У випадку із моделлю прогнозу параметра заповнюваності ситуація буде дещо іншою. Оскільки заповнюваність обчислюється за формулою (3.1), то можна зробити логічний висновок, що ті параметри які включені до цієї формули потрібно виключити із розгляду при аналізі значущості, бо очевидно що заповнюваність буде найбільшою мірою залежати саме від них. Тобто потрібно видалити параметри «matched ad requests» та «ad requests». Після цього буде виконуватись аналіз значущості 14 параметрів відносно метрики заповнюваності, бо інформаційні атрибути апріорі відкидаються.

Власне сам аналіз значущості було вирішено проводити за допомогою двох методів, а саме – із використанням кореляційної матриці та із застосуванням методу градієнтного бустингу, що адаптований для визначення важливості параметрів моделі. Кореляційний аналіз буде проведений, оскільки це типова процедура вибору значущих параметрів, але вона дає результати тільки стосовно лінійних залежностей параметрів. Саме тому, наступним кроком буде використання методу XGBoost для визначення важливості тих чи інших параметрів. При цьому даний метод враховує всі види залежностей між досліджуваними параметрами. Після того, як будуть визначені найбільш важливі параметри, можна переходити безпосередньо до моделювання. І на цьому етапі за допомогою варіювання кількості параметрів, що розглядаються, можна змінювати складність та точність конкретної моделі.

Наступним етапом буде власне моделювання. Було вирішено скористатись наступною схемою. Спочатку потрібно побудувати модель множинної регресії на основі наявних параметрів. При виконанні цієї процедури можна врахувати попередній крок, тобто вибирати параметри, які будуть включені в модель, залежно від їхньої значущості. Таким чином можна варіювати складність побудованої моделі, а також її точність. Тому є можливість підібрати підходящу для конкретного випадку модель. Після того, як модель множинної регресії для вхідних статистичних даних буде побудована, потрібно знайти змодельоване значення шуканої величини. Для різних моделей це буде дохід або параметр заповнюваності. На основі отриманого часового ряду будуть вже будуватись подальші моделі прогнозу обраних статистичних параметрів.

Після побудови множинної регресії та перед власне побудовою моделей прогнозу, потрібно провести процедуру розбиття вихідної вибірки на навчальну тестову. При цьому побудову моделей здійснювати на навчальні вибірці, а виконувати прогноз на тестовій. Враховуючи кількість наявних даних та період прогнозу, було вирішено розбити вихідний набір даних у співвідношення 202 / 10. Після побудови прогнозу можна буде використати тестовий набір

даних для перевірки точності прогнозу. Після виконання наведених вище маніпуляцій можна побудувати прогноз «наперед».

Для того, щоб побудувати найбільш підходящу модель для конкретного випадку, було вирішено використати процедуру поступового ускладнення моделі. Тобто спочатку будується певна проста модель, і на її основі пізніше отримуємо більш складну модель, що потенційно повинна володіти більшою точністю. За допомогою такого ітеративного процесу планується побудувати таку модель, яка найкраще буде виконувати покладені на неї задачі.

Таким чином, враховуючи все сказане вище, у даній роботі буде реалізована наступна послідовність кроків:

- побудова множинної регресії;
- побудова моделі *AR*;
- побудова моделі *ARMA*;
- побудова моделі *ARIMA*;
- побудова моделі *ARIMAX*.

Далі на основі статистичних критеріїв буде обрано найкращу модель для подальшого прогнозування. При цьому, варто скористатися не лише похибками, але також інформаційними критеріями для порівняння всіх побудованих моделей.

Таким чином, наступна задача полягає у побудові прогнозу «наперед». Можна побудувати декілька моделей прогноз, наприклад, на 1 тиждень та 2 тижні.

Усі наведені вище маніпуляції потрібно здійснити двічі, оскільки нам потрібно побудувати моделі прогнозу як для доходу, так і для параметру заповнюваності. Тобто в результаті ми повинні отримати дві моделі прогнозу для різних статистичних характеристик рекламного трафіку. Отримані моделі будуть характеризуватись середнім значенням отриманого прогнозу, а також двома інтервалами прогнозу. А саме – 80% та 95% інтервалами. Отримані прогнози доходу та показника заповнюваності, власне кажучи, і є кінцевою метою даної роботи.



### 3.1.3 Блок-схема розв'язання поставленої задачі

Після того, як були сформульовані конкретні задачі, які необхідно розв'язати для здійснення моделювання, потрібно розробити блок-схему розв'язання поставленої задачі прогнозування доходу та заповнюваності. Розроблена блок-схема представлена на рисунку 3.1.

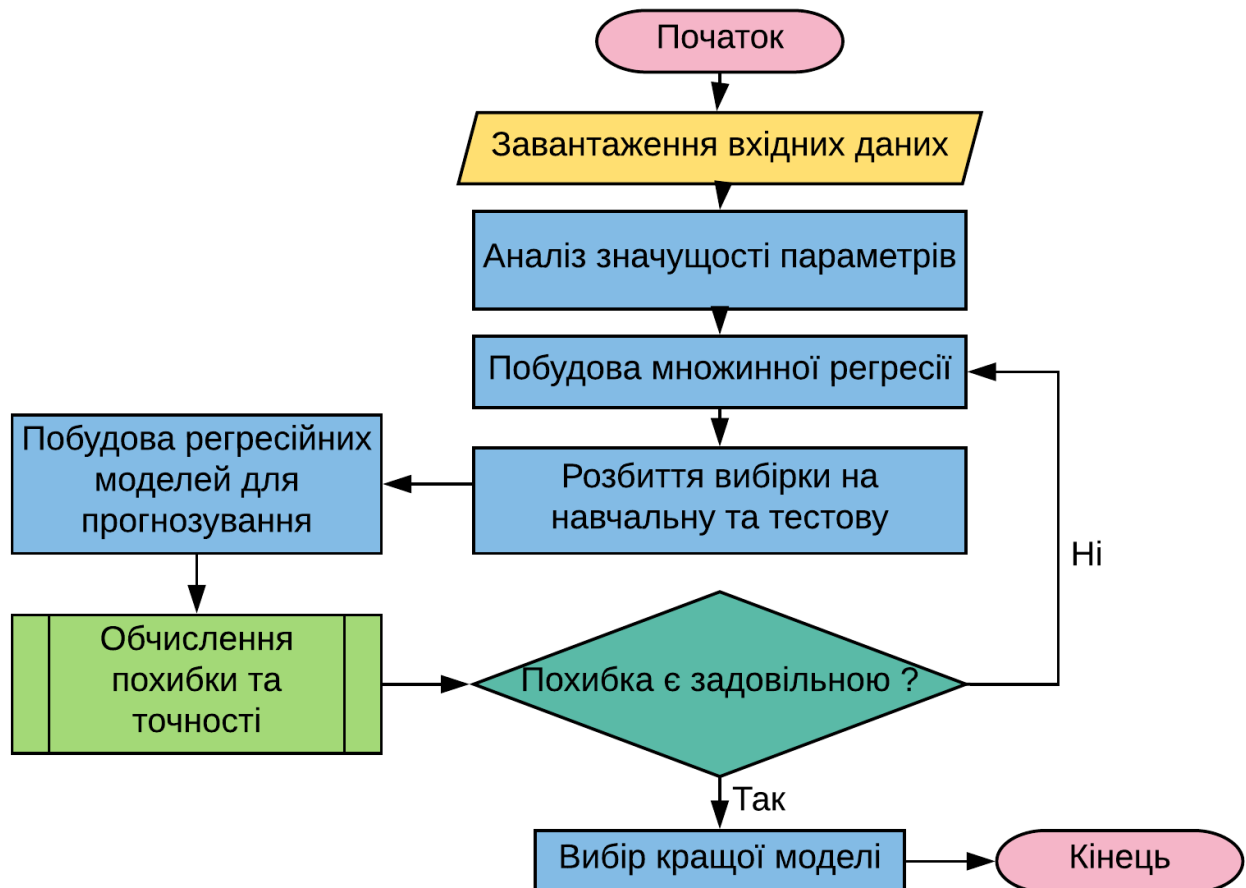


Рисунок 3.1 – Блок-схема розв'язання задачі прогнозування доходу та заповнюваності

### 3.2 Попередня обробка вхідних даних

Спочатку наведемо фрагмент вхідного набору даних, що містить перші 15 записів та перші 7 атрибутів цього набору. Даний фрагмент в узагальненому вигляді описує вхідні дані, та представлений на рисунку 3.2.

date	ad_tag_name	ad_requests	matched_ad_requests	ad_requests_coverage	clicks	ad_requests_ctr
2017-11-14	site 1	59642	504	0,0085	20	0,0003
2017-11-14	site 2	67532	1241	0,0184	71	0,0011
2017-11-14	site 3	16057	0	0	0	0
2017-11-14	site 4	5059	2013	0,3979	45	0,0089
2017-11-14	site 5	656011	86216	0,1314	312	0,0005
2017-11-14	site 6	618764	137218	0,2218	371	0,0006
2017-11-14	site 7	123792	49190	0,3974	1528	0,0123
2017-11-14	site 8	104987	67980	0,6475	3164	0,0301
2017-11-14	site 9	433339	208246	0,4806	2589	0,006
2017-11-14	site 10	54687	11603	0,2122	338	0,0062
2017-11-14	site 11	13903	2012	0,1447	21	0,0015
2017-11-14	site 12	16881	15929	0,9436	46	0,0027
2017-11-14	site 13	12483	9296	0,7447	399	0,032
2017-11-14	site 14	27329	18880	0,6908	1098	0,0402
2017-11-14	site 15	112239	47556	0,4237	257	0,0023

Рисунок 3.2 – Початковий фрагмент вхідного набору даних

Перед тим як приступити до моделювання, потрібно виконати декілька важливих етапів. А саме, спочатку потрібно відповідним чином підгодувати вхідні дані. Зокрема, потрібно зазначити, що набір даних, який був взятий до аналізу дещо відрізняється від того, який нам потрібен для виконання моделювання. Опишемо проблему більш детально. У початковому наборі даних є 19 параметрів, із яких 2 інформаційні параметри та 17 числових параметрів, що являють собою статистичні метрики для трафіка цифрової реклами. При цьому, варто зазначити, що для виконання прогнозування потрібний агрегований набір даних по дням. Тобто треба виконати агрегацію вихідного набору даних по дням, при цьому параметри, що являють собою кількість чогось, потрібно агрегувати за допомогою суми, а ті параметри, що є певним відсотковим відношенням, або, як винятки: ціна за клік та ціна за тисячу показів рекламного оголошення, потрібно агрегувати за допомогою середнього значення. Оскільки початкове розбиття вихідного набору даних було здійснене за параметрами дати та назви рекламного тегу, а в результаті нам потрібно отримати агреговані дані по дням, то відбудеться агрегація статистичного набору даних по рекламних тегах для кожного дня. Тип агрегації був вказаний вище. При цьому, після агрегації інформаційний параметр «назва рекламного тегу» зникне, і у нас залишиться лише один інформаційний параметр – це дата показу рекламного оголошення.

Також не потрібно забувати про такі важливі етапи підготовки статистичних даних до аналізу як очищення даних та приведення типів. Зокрема, спочатку потрібно провести процедуру приведення типів даних статистичних метрик. Вона полягає у тому, щоб для відповідних параметрів були підбрані найбільш підходящі типи даних. Тобто для параметрів, які характеризують кількість чогось, потрібно використати цілочисельний тип даних, а для параметрів що являють собою певне співвідношення або деяку ціну – потрібно використати тип даних для дійсних чисел, тобто із десятковою крапкою. Окремо можна згадати про параметри «дата» та «назва рекламного тегу». Для дати потрібно використати тип даних, що характеризує дату, для назви рекламного тегу – текстовий тип даних. Після того, як описані вище процедури будуть виконані, потрібно приступити до очищення даних. А саме – потрібно обробити порожні значення. У нашому випадку ми будемо замінювати порожні значення на «0», оскільки це буде означати, що даний рекламний тег є неактивним, про що власне кажучи і свідчить порожнє значення для певного статистичного параметру рекламного тегу. Після цього потрібно виконати очищення набору даних від так званих «викидів», тобто ймовірно помилкових значень, що були записані в поле певної метрики. У нашому випадку було виявлено наступні «викиди»: значення статистичних метрик, які менші від нуля. Ці значення є помилковими, оскільки всі метрики статистичного рекламного трафіку цифрової реклами апріорі є невід’ємними. Тому такі значення потрібно виявляти ще до проведення моделювання, оскільки вони можуть привести до неправильних висновків у процесі моделювання. Таким чином, було вирішено замінювати такі значення також на «0». Це досить логічно, оскільки помилкові значення свідчать про те, що даний тег насправді не використовувався. Тому заміна на значення «0» виглядає досить природньо, вона не змінює значення даних, і якраз свідчить про те, що рекламний тег не є активним, що привело нас до попереднього пункту. Після того, як усі ці підготовчі етапи статистичних даних буде виконано, можна переходити до безпосередньо моделювання.

### 3.3 Вибір значущих параметрів

На даному етапі потрібно оцінити важливість наявних параметрів, перед тим, як приступати безпосередньо до моделювання. Варто зазначити, що у вихідному наборі даних кількість параметрів є досить великою, і на практиці для побудови моделей це не завжди добре. Тому досить логічно на наступному кроці виконати аналіз значущості параметрів. Оскільки у даній роботі виконується побудова двох моделей прогнозу, то потрібно також проводити аналіз значущості параметрів окремо для кожного випадку. Як було зазначено вище, ми здійснили агрегацію вихідного набору даних, і тепер залишилось вже не 19 параметрів, а 18, тому що атрибут «назва рекламного тегу» зник в результаті агрегування даних по дням. Тепер опишемо як буде виконуватись аналіз значущості наявних параметрів. Для моделі прогнозу доходу будемо діяти наступним чином. В агрегованому набору даних залишилось 17 числових статистичних характеристик та 1 інформаційний атрибут. При цьому, із цих 17 параметрів один параметр – власне сам дохід. Тому будемо в даному випадку буде перевірятись значущість 16 параметрів по відношенню до доходу, оскільки інформаційні атрибути у даній процедурі відкидаються. У випадку із моделлю прогнозу параметра заповнюваності ситуація буде дещо іншою. Оскільки заповнюваність обчислюється за формулою (3.1), то можна зробити логічний висновок, що ті параметри які включені до цієї формули потрібно виключити із розгляду при аналізі значущості, бо очевидно що заповнюваність буде найбільшою мірою залежати саме від них. Тобто потрібно видалити параметри «matched ad requests» та «ad requests». Після цього буде виконуватись аналіз значущості 14 параметрів відносно метрики заповнюваності, бо інформаційні атрибути апріорі відкидаються.

Власне сам аналіз значущості було вирішено проводити за допомогою двох методів, а саме – із використанням кореляційної матриці та із застосуванням методу градієнтного бустингу, що адаптований для визначення важливості параметрів моделі. Кореляційний аналіз буде проведений, оскільки це типова

процедура вибору значущих параметрів, але вона дає результати тільки стосовно лінійних залежностей параметрів. Саме тому, наступним кроком буде використання методу XGBoost для визначення важливості тих чи інших параметрів [27]. При цьому даний метод враховує всі види залежностей між досліджуваними параметрами. Після того, як будуть визначені найбільш важливі параметри, можна переходити безпосередньо до моделювання. І на цьому етапі за допомогою варіювання кількості параметрів, що розглядаються, можна змінювати складність та точність конкретної моделі.

Тепер в таблицях 3.1-3.4 наведемо результати проведеного аналізу.

Таблиця 3.1 – Коефіцієнти кореляції для показника заповнюваності

<b>parameters</b>	<b><i>ad_requests_coverage</i></b>
<b><i>ad_requests</i></b>	
<b><i>matched_ad_requests</i></b>	
<b><i>clicks</i></b>	0,313279239
<b><i>earnings</i></b>	0,55899458
<b><i>ad_impressions</i></b>	0,486534831
<b><i>video_ad_skippable_views</i></b>	0,426867821
<b><i>active_view_eligible_count</i></b>	0,486598543
<b><i>active_view_measurable_count</i></b>	0,485519065
<b><i>active_view_viewable_count</i></b>	0,454520146
<b><i>ad_requests_coverage</i></b>	
<b><i>ad_requests_ctr</i></b>	0,000963579
<b><i>matched_ad_requests_ctr</i></b>	0,432421831
<b><i>individual_ad_impressions_ctr</i></b>	0,413444552
<b><i>cost_per_click</i></b>	0,214325317
<b><i>ad_impressions_rpm</i></b>	0,111541263
<b><i>video_ad_skippable_skip_ratio</i></b>	0,267885451
<b><i>video_ad_skippable_vtr</i></b>	0,212886662

Таблиця 3.2 – Коефіцієнти кореляції для доходу

<b>parameters</b>	<b><i>earnings</i></b>
<b><i>ad_requests</i></b>	0,181422986
<b><i>matched_ad_requests</i></b>	0,092267193
<b><i>clicks</i></b>	0,006913077
<b><i>earnings</i></b>	

Продовження таблиці 3.2

<b>parameters</b>	<b>earnings</b>
<b><i>ad_impressions</i></b>	0,02310064
<b><i>video_ad_skippable_views</i></b>	0,520574827
<b><i>active_view_eligible_count</i></b>	0,023229263
<b><i>active_view_measurable_count</i></b>	0,022512074
<b><i>active_view_viewable_count</i></b>	0,013358076
<b><i>ad_requests_coverage</i></b>	0,55899458
<b><i>ad_requests_ctr</i></b>	0,031654862
<b><i>matched_ad_requests_ctr</i></b>	0,221559917
<b><i>individual_ad_impressions_ctr</i></b>	0,197329556
<b><i>cost_per_click</i></b>	0,201062948
<b><i>ad_impressions_rpm</i></b>	0,020299547
<b><i>video_ad_skippable_skip_ratio</i></b>	0,182436327
<b><i>video_ad_skippable_vtr</i></b>	0,014653604

Таблиця 3.3 – Важливість параметрів по відношенню до показника заповнюваності, що отримана методом XGBoost

<b>parameters</b>	<b><i>ad_requests_coverage</i></b>
<b><i>active_view_eligible_count</i></b>	0,034755134
<b><i>active_view_measurable_count</i></b>	0,037914692
<b><i>active_view_viewable_count</i></b>	0,060031596
<b><i>ad_impressions</i></b>	0,066350711
<b><i>ad_impressions_rpm</i></b>	0,078988942
<b><i>ad_requests</i></b>	
<b><i>ad_requests_coverage</i></b>	
<b><i>ad_requests_ctr</i></b>	0,078988942
<b><i>clicks</i></b>	0,088467615
<b><i>cost_per_click</i></b>	0,060031596
<b><i>earnings</i></b>	0,088467615
<b><i>individual_ad_impressions_ctr</i></b>	0,090047393
<b><i>matched_ad_requests</i></b>	
<b><i>matched_ad_requests_ctr</i></b>	0,09478673
<b><i>video_ad_skippable_skip_ratio</i></b>	0,071090047
<b><i>video_ad_skippable_views</i></b>	0,09478673
<b><i>video_ad_skippable_vtr</i></b>	0,055292259

Таблиця 3.4 – Важливість параметрів по відношенню до доходу, що отримана методом XGBoost

<b>parameters</b>	<b>earnings</b>
<i>active_view_eligible_count</i>	0,05952381
<i>active_view_measurable_count</i>	0,034461153
<i>active_view_viewable_count</i>	0,036967419
<i>ad_impressions</i>	0,061403509
<i>ad_impressions_rpm</i>	0,047619048
<i>ad_requests</i>	0,079573935
<i>ad_requests_coverage</i>	0,073934837
<i>ad_requests_ctr</i>	0,062030075
<i>clicks</i>	0,072681704
<i>cost_per_click</i>	0,05952381
<i>earnings</i>	
<i>individual_ad_impressions_ctr</i>	0,075814536
<i>matched_ad_requests</i>	0,067669173
<i>matched_ad_requests_ctr</i>	0,060150376
<i>video_ad_skippable_skip_ratio</i>	0,068922306
<i>video_ad_skippable_views</i>	0,077067669
<i>video_ad_skippable_vtr</i>	0,062656642

### 3.4 Розбиття вихідної вибірки на навчальну та тестову

З метою перевірки точності побудованих моделей, було вирішено розбивати вихідну вибірку на дві частини, а саме – навчальну та тестову. Значення при розбитті були вибрані досить стандартними – 202 на навчальну та 10 на тестову. Тобто ми використовуємо для перевірки моделей 10 значень із вихідної вибірки. По суті це аналог середньострокового прогнозу. Після перевірки побудованих моделей на тестовій вибірці, можна будувати прогнози «наперед», причому модель потрібно будувати по всім наявним даним [28].

Наведемо графічні представлення часових рядів для параметрів доходу та заповнюваності. А також зобразимо їх розбиття на навчальну та тестову частини.

Результати розбиття наведемо на рисунках 3.3, 3.4.

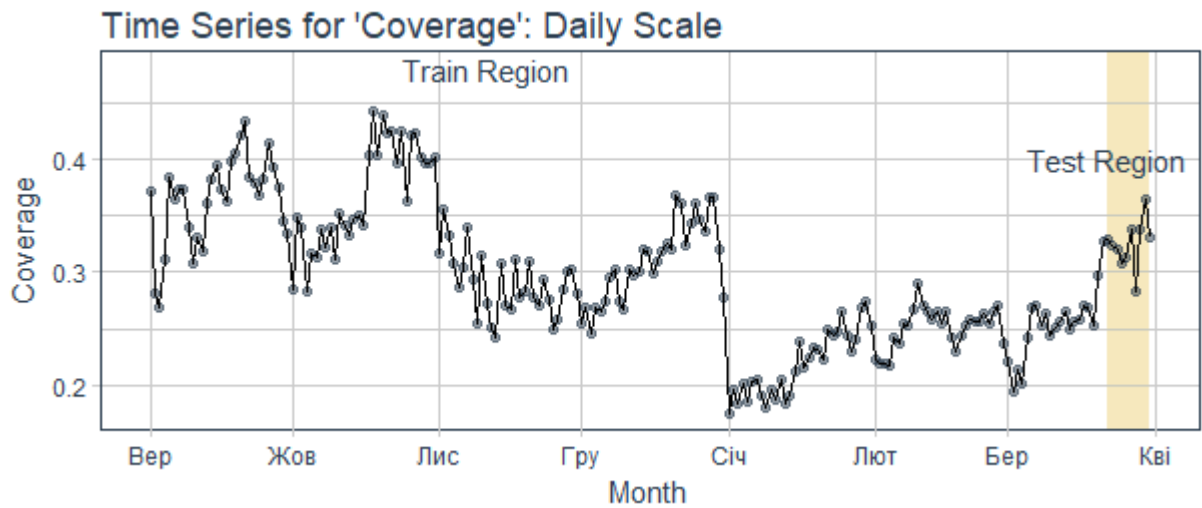


Рисунок 3.3 – Графічне представлення часового ряду для заповнюваності, а також розбиття на навчальну та тестову вибірки

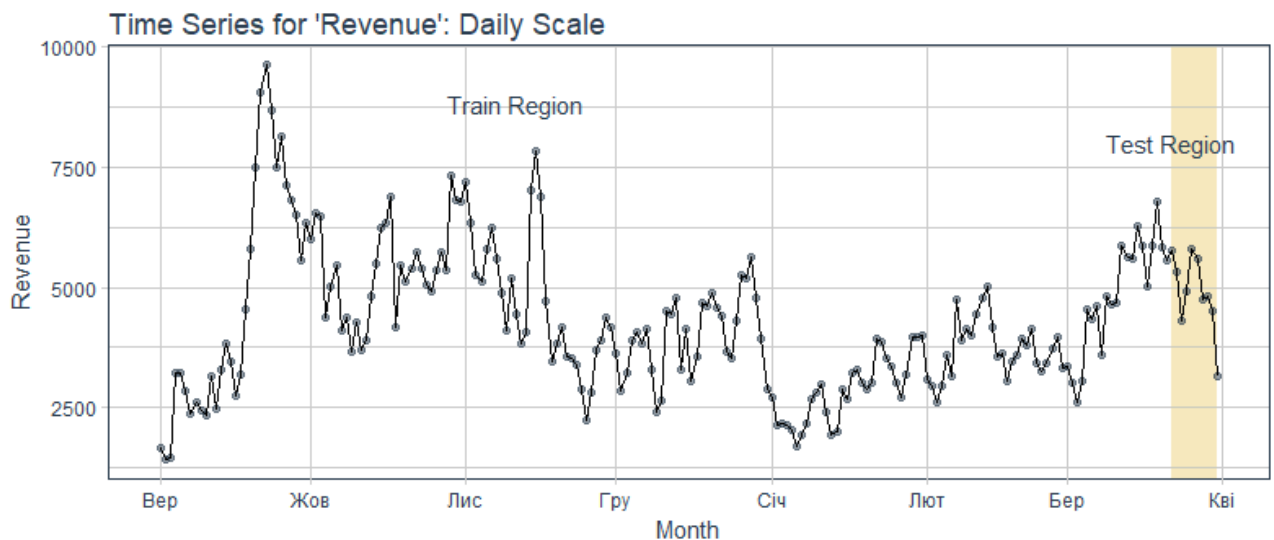


Рисунок 3.4 – Графічне представлення часового ряду для доходу, а також розбиття на навчальну та тестову вибірки

### 3.5 Побудова моделей прогнозу доходу та заповнюваності

#### 3.5.1 Побудова моделі множинної регресії

На даному етапі потрібно виконати побудову моделей множинної регресії для двох випадків – прогнозу доходу та прогнозу параметра заповнюваності.

У випадку із доходом, ми будемо будувати модель на основі всіх числових параметрів, за виключенням самого доходу. А при моделюванні заповнюваності потрібно видалити не лише сам параметр, а ще й метрики для кількості запитів,



а також для кількості зіставлених запитів, оскільки заповнюваність напряму залежить від цих метрик, тому що саме так вона і вираховується.

Наведемо на рисунках 3.5-3.8 коефіцієнти отримані при моделюванні та графічне представлення часових рядів, отриманих на основі побудованих моделей.

```
Residuals:
    Min       1Q   Median       3Q      Max
-1571.5  -417.1   -79.4    401.8   3631.6

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -6.874e+03  1.358e+03  -5.062 9.58e-07 ***
ad_requests  9.037e-05  2.454e-05   3.683 0.000298 ***
matched_ad_requests 2.308e-03  3.151e-04   7.326 6.14e-12 ***
clicks       8.406e-04  1.440e-03   0.584 0.560077
ad_impressions -9.377e-03  9.668e-03  -0.970 0.333320
video_ad_skippable_views -6.305e-04  2.006e-03  -0.314 0.753569
active_view_eligible_count 6.660e-03  1.019e-02   0.654 0.514142
active_view_measurable_count 6.776e-04  3.179e-03   0.213 0.831417
active_view_viewable_count -9.326e-04  3.429e-04  -2.719 0.007129 **
ad_requests_coverage 7.609e+03  1.929e+03   3.945 0.000111 ***
ad_requests_ctr -2.959e+04  3.106e+04  -0.953 0.341933
matched_ad_requests_ctr 1.059e+05  3.742e+04   2.829 0.005157 **
individual_ad_impressions_ctr -7.035e+04  1.916e+04  -3.671 0.000312 ***
cost_per_click 2.094e+01  3.417e+02   0.061 0.951206
ad_impressions_rpm 1.215e+03  2.053e+02   5.917 1.45e-08 ***
video_ad_skippable_skip_ratio 8.857e+03  3.065e+03   2.890 0.004295 **
video_ad_skippable_vtr -7.553e+03  5.304e+03  -1.424 0.156081
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 724.5 on 195 degrees of freedom
Multiple R-squared:  0.7839,    Adjusted R-squared:  0.7662
F-statistic: 44.21 on 16 and 195 DF,  p-value: < 2.2e-16
```

Рисунок 3.5 – Модель множинної регресії для доходу

```
Residuals:
    Min       1Q   Median       3Q      Max
-0.07776 -0.01933 -0.00228  0.01832  0.10078

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.615e-01  5.401e-02   4.843 2.58e-06 ***
clicks       -8.561e-08  5.995e-08  -1.428 0.154908
earnings      9.233e-06  2.086e-06   4.426 1.59e-05 ***
ad_impressions 2.148e-07  4.075e-07   0.527 0.598615
video_ad_skippable_views 4.771e-07  6.994e-08   6.822 1.08e-10 ***
active_view_eligible_count -1.203e-07  4.286e-07  -0.281 0.779261
active_view_measurable_count -9.687e-08  1.316e-07  -0.736 0.462565
active_view_viewable_count -3.540e-08  1.397e-08  -2.535 0.012029 *
ad_requests_ctr 9.119e+00  1.144e+00   7.971 1.25e-13 ***
matched_ad_requests_ctr -8.229e+00  1.428e+00  -5.760 3.19e-08 ***
individual_ad_impressions_ctr 3.238e+00  7.529e-01   4.300 2.68e-05 ***
cost_per_click 6.262e-02  1.338e-02   4.682 5.28e-06 ***
ad_impressions_rpm -4.075e-02  8.397e-03  -4.853 2.47e-06 ***
video_ad_skippable_skip_ratio 4.671e-01  1.272e-01   3.674 0.000308 ***
video_ad_skippable_vtr -4.863e-01  2.118e-01  -2.296 0.022721 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03073 on 197 degrees of freedom
Multiple R-squared:  0.7726,    Adjusted R-squared:  0.7564
F-statistic: 47.8 on 14 and 197 DF,  p-value: < 2.2e-16
```

Рисунок 3.6 – Модель множинної регресії для параметра заповнюваності

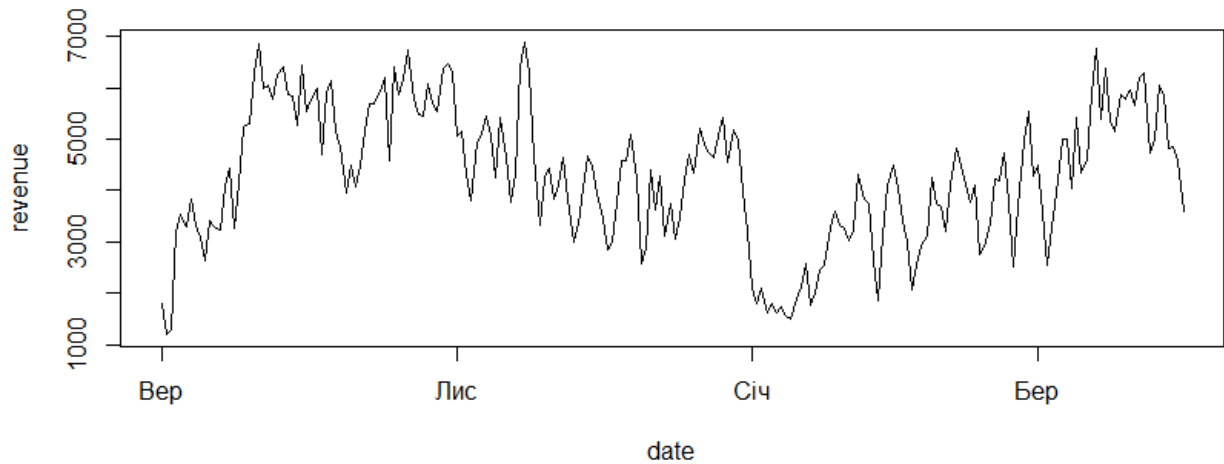


Рисунок 3.7 – Графічне представлення часового ряду для змодельованого доходу

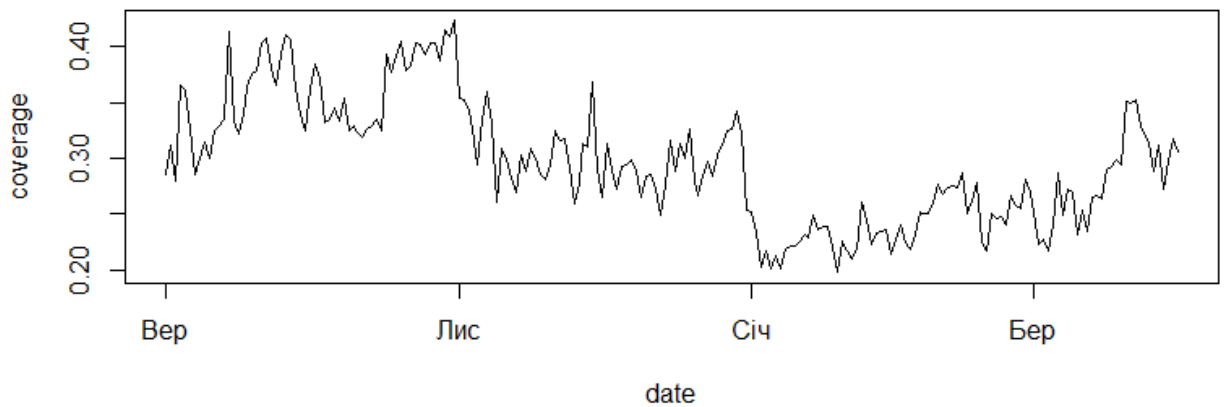


Рисунок 3.8 – Графічне представлення часового ряду для змодельованого параметру заповнюваності

### 3.5.2 Побудова моделі AR

Тепер перейдемо до наступного кроку, а саме – до побудови моделей  $AR(p)$ . Як завжди буде виконана побудова двох моделей – для доходу та заповнюваності. Після цього наведемо значення показників адекватності, інформативності моделей, а також залишки моделей на навчальній вибірці та значення похибок і прогноз та тестовій вибірці [29].

Для заповнюваності отримано модель  $AR(3)$ , а для доходу –  $AR(1)$ .

На рисунках 3.9-3.16 зображено результати моделювання.

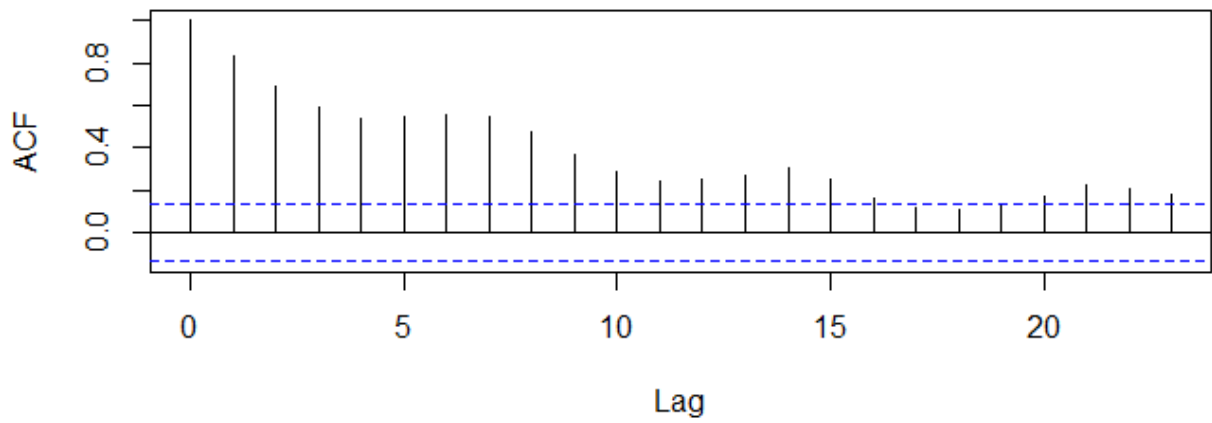


Рисунок 3.9 – АКФ для змодельованого параметру доходу

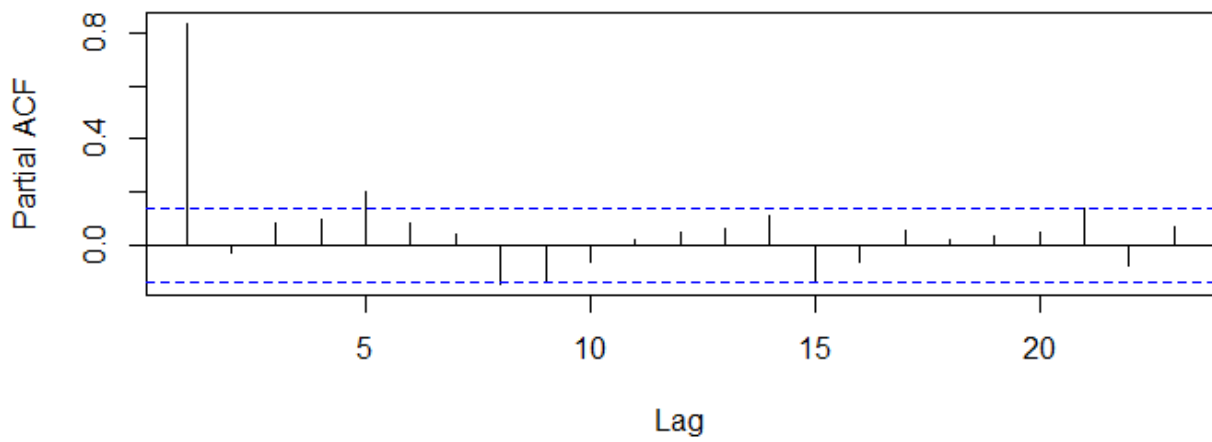


Рисунок 3.10 – ЧАКФ для змодельованого параметру доходу

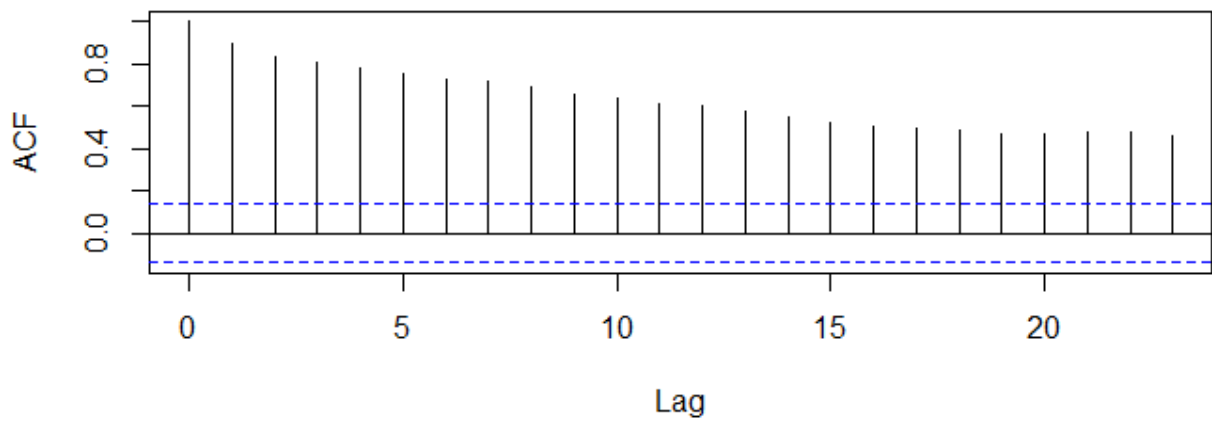


Рисунок 3.11 – АКФ для змодельованого параметру заповнюваності

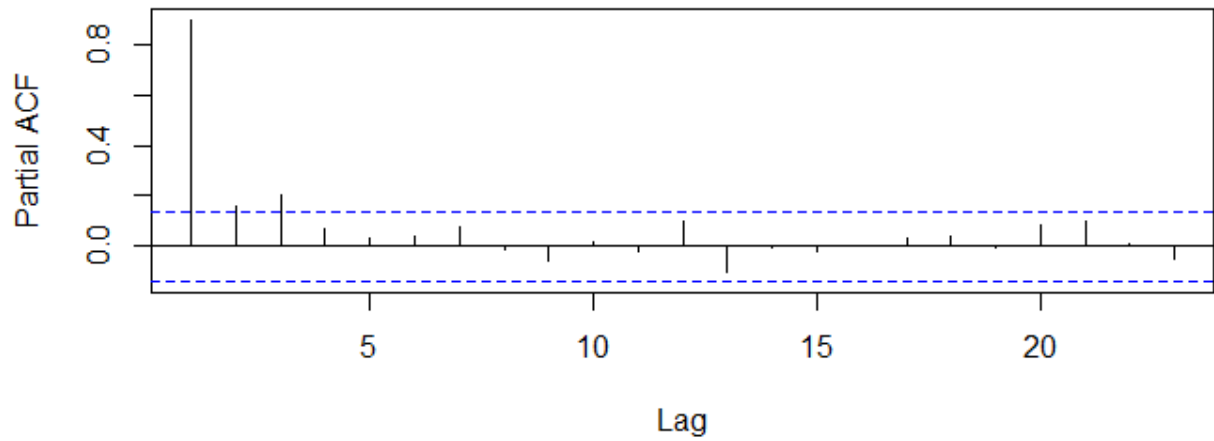


Рисунок 3.12 – ЧАКФ для змодельованого параметру заповнюваності

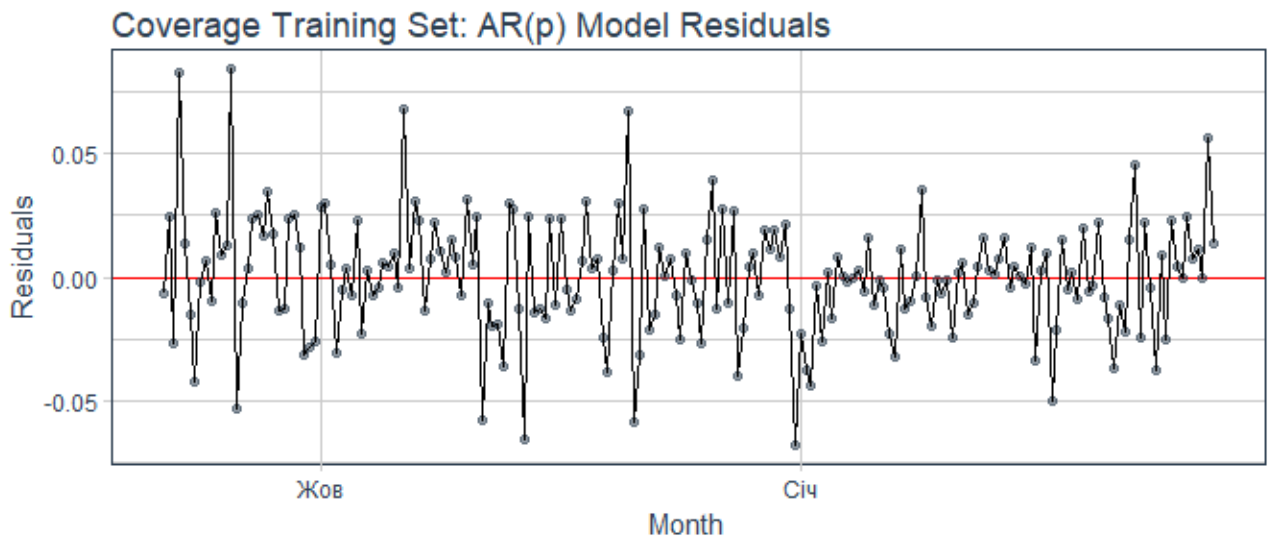


Рисунок 3.13 – Залишки для моделі AR для заповнюваності

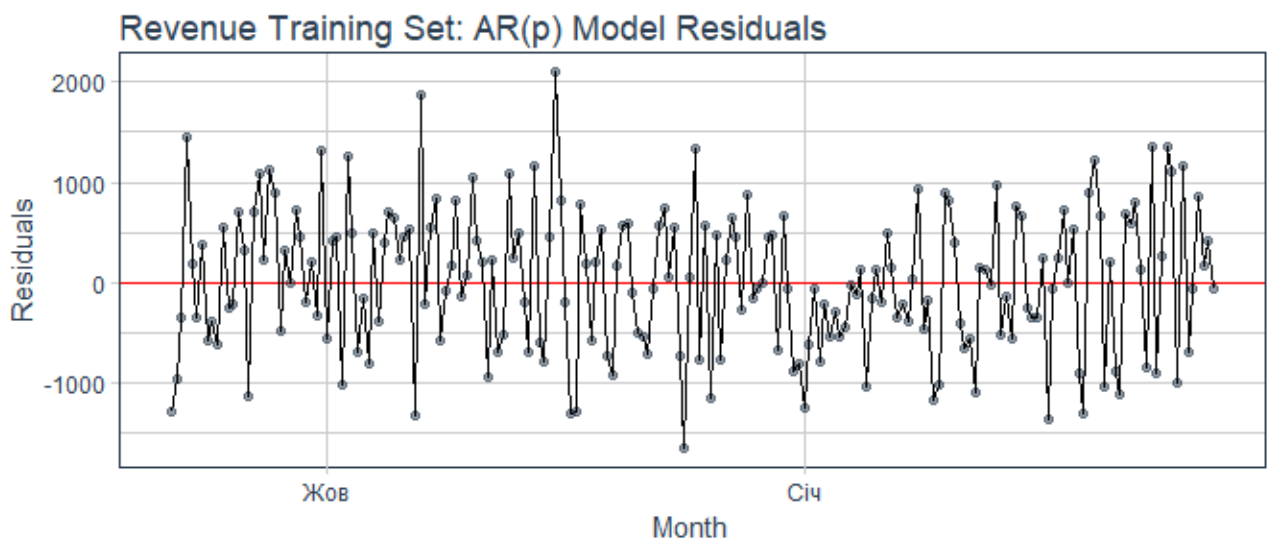


Рисунок 3.14 – Залишки для моделі AR для доходу

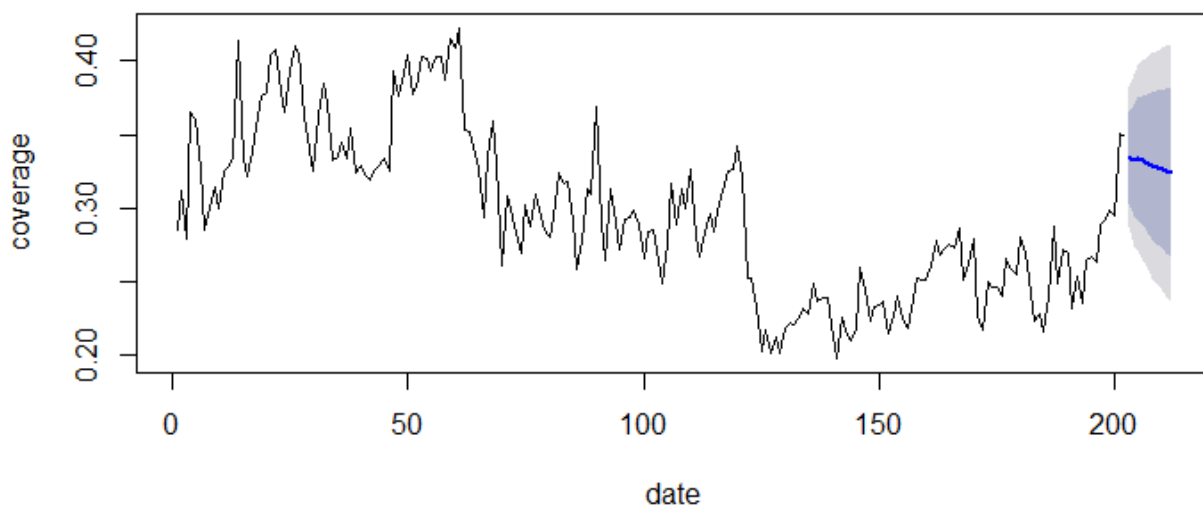


Рисунок 3.15 – Прогноз на тестовій вибірці за моделлю AR для заповнюваності

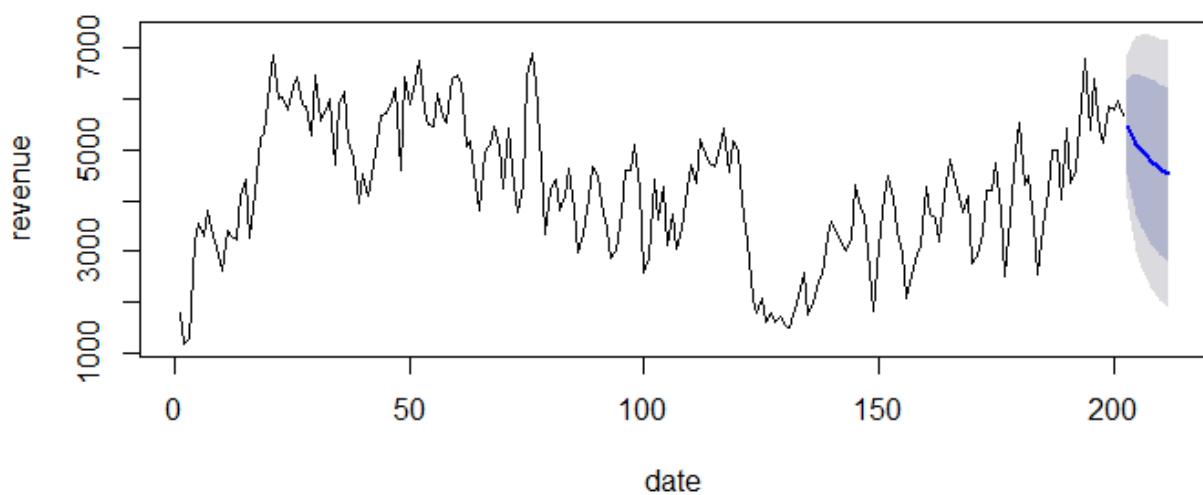


Рисунок 3.16 – Прогноз на тестовій вибірці за моделлю AR для доходу

У таблиці 3.5 наведено результати обчислення критеріїв, а в таблиці 3.6 наведено результати обчислення похибок.

Таблиця 3.5 – Значення критеріїв для доходу та заповнюваності

Критерій	Дохід	Заповнюваність
<i>AIC</i>	3233.09	-924.14
<i>BIC</i>	3243.01	-904.29
$R^2$	0.714015	0.8180255
<i>DW</i>	1.981992	1.993624

Таблиця 3.6 – Значення похибок для доходу та заповнюваності

Похибка	Дохід	Заповнюваність
<i>RMSE</i>	710.3305	0.02374302
<i>MAE</i>	587.1062	0.01780827
<i>MAPE</i>	15.69017	6.044972
<i>MASE</i>	0.9687749	0.9460523

### 3.5.3 Побудова моделі ARMA

Перейдемо до побудови моделей ARMA. Як завжди буде виконана побудова двох моделей – для доходу та заповнюваності. Після цього наведемо значення показників адекватності, інформативності моделей, а також залишки моделей на навчальній вибірці та значення похибок і прогноз та тестовій вибірці.

Для доходу отримано модель ARMA(1, 0), а для заповнюваності – ARMA(2, 3).

Наведемо результати моделювання на рисунках 3.17-3.20.

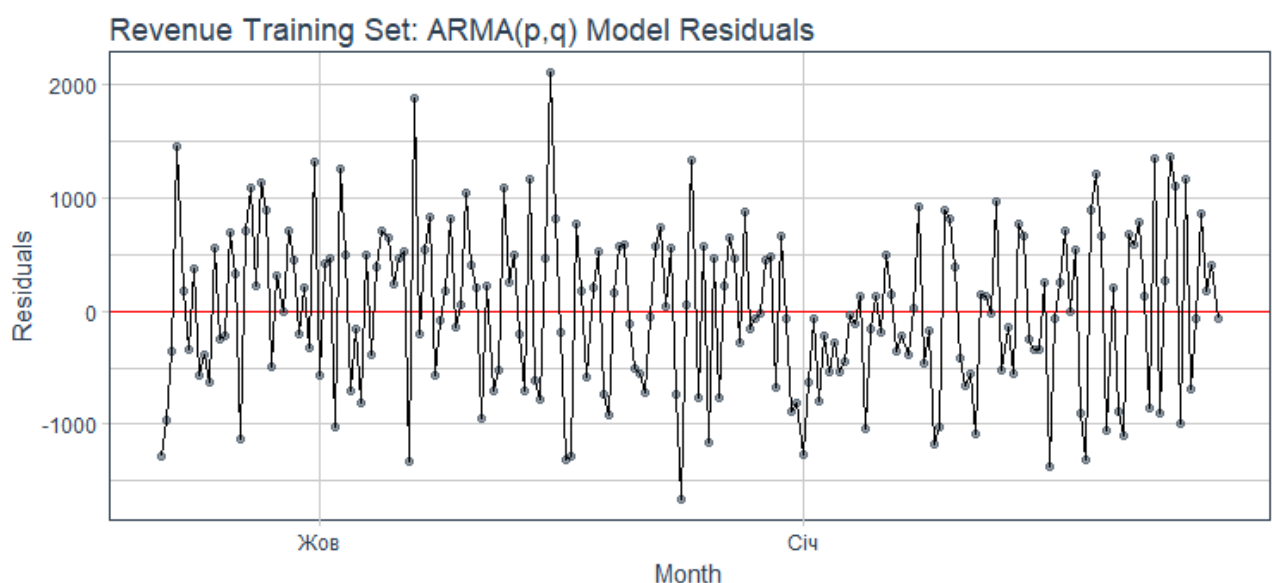


Рисунок 3.17 – Залишки для моделі ARMA для доходу

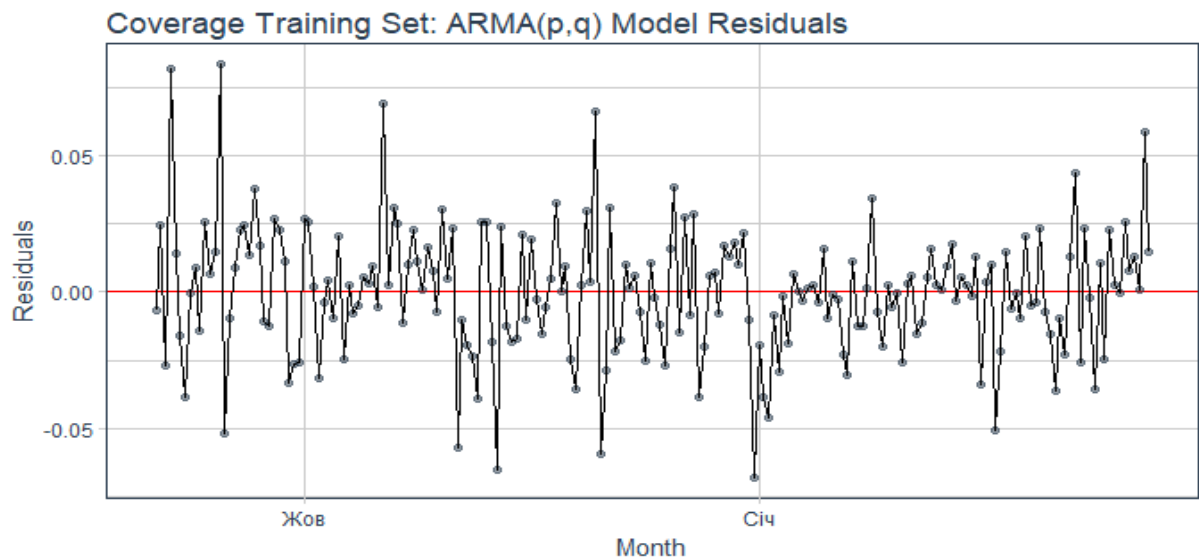


Рисунок 3.18 – Залишки для моделі ARMA для заповнюваності

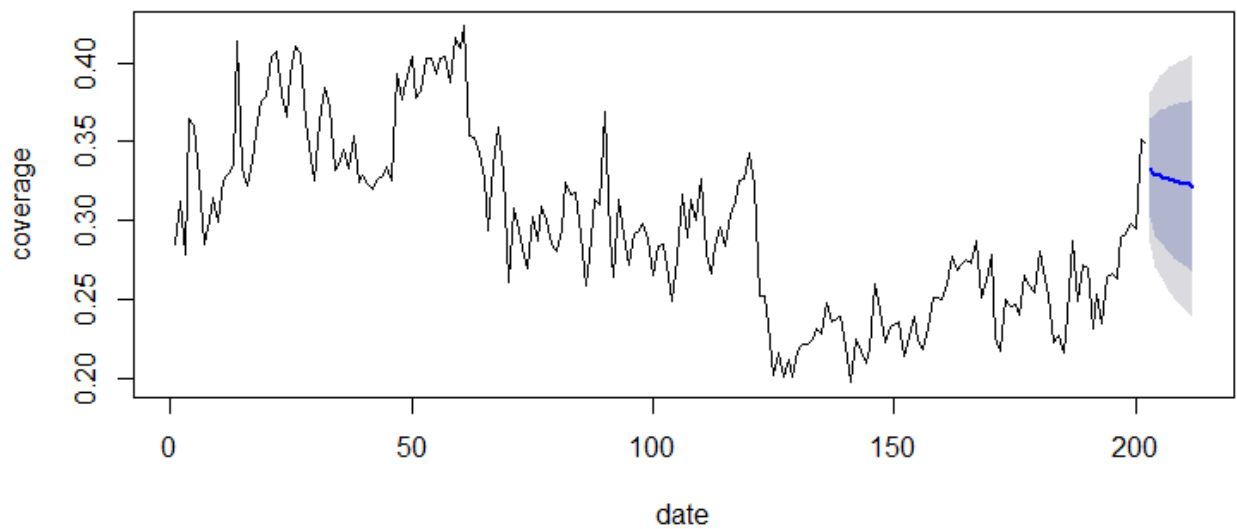


Рисунок 3.19 – Прогноз на тестовій вибірці за моделлю ARMA для заповнюваності

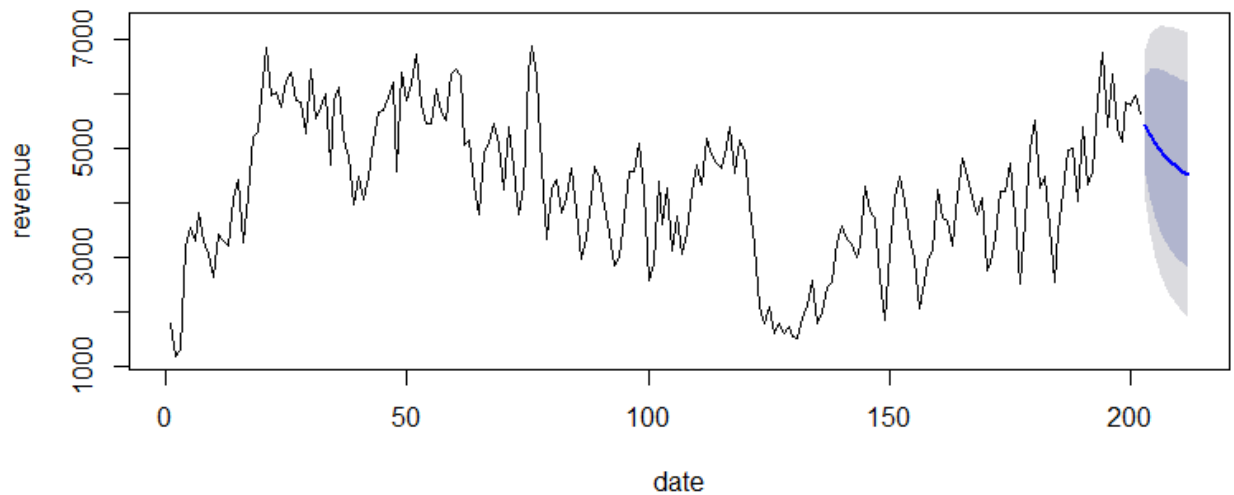


Рисунок 3.20 – Прогноз на тестовій вибірці за моделлю ARMA для доходу

У таблиці 3.7 наведено результати обчислення критеріїв, а в таблиці 3.8 наведено результати обчислення похибок.

Таблиця 3.7 – Значення критеріїв для доходу та заповнюваності

<b>Критерій</b>	<b>Дохід</b>	<b>Заповнюваність</b>
<i>AIC</i>	3233.09	-922.83
<i>BIC</i>	3243.01	-899.68
<i>R<sup>2</sup></i>	0.714015	0.8186622
<i>DW</i>	1.981992	2.003657

Таблиця 3.8 – Значення похибок для доходу та заповнюваності

<b>Похибка</b>	<b>Дохід</b>	<b>Заповнюваність</b>
<i>RMSE</i>	710.3305	0.02370153
<i>MAE</i>	587.1062	0.0177796
<i>MAPE</i>	15.69017	6.045356
<i>MASE</i>	0.9687749	0.9445291

### 3.5.4 Побудова моделі ARIMA

Наступним етапом буде побудова моделі ARIMA. Як завжди буде виконана побудова двох моделей – для доходу та заповнюваності. Після цього наведемо значення показників адекватності, інформативності моделей, а також залишки моделей на навчальній вибірці та значення похибок і прогноз та тестовій вибірці.

У даному випадку отримали наступні результати. Для заповнюваності було побудовано модель ARIMA(1, 1, 3), для доходу – ARIMA(2, 1, 3).

Результати моделювання наведемо на рисунках 3.21-3.24.



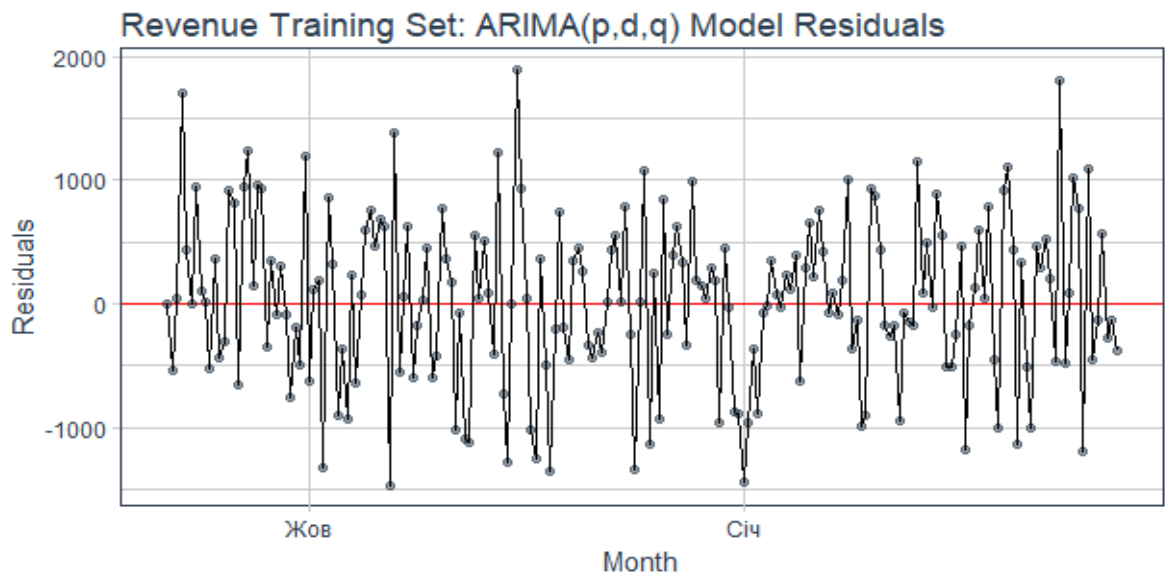


Рисунок 3.21 – Залишки для моделі ARIMA для доходу

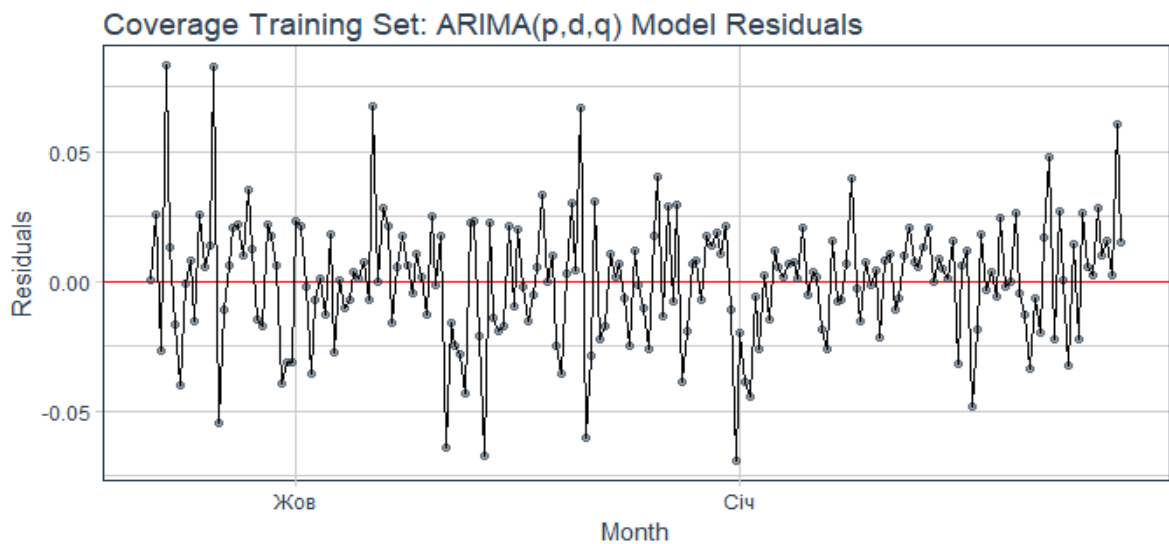


Рисунок 3.22 – Залишки для моделі ARIMA для заповнюваності

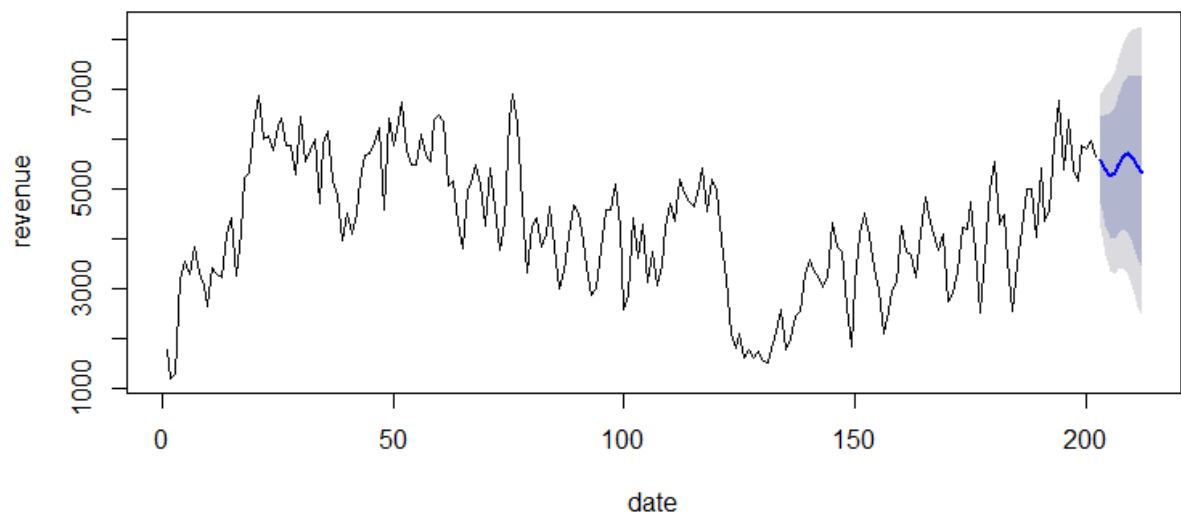


Рисунок 3.23 – Прогноз на тестовій вибірці за моделлю ARIMA для доходу

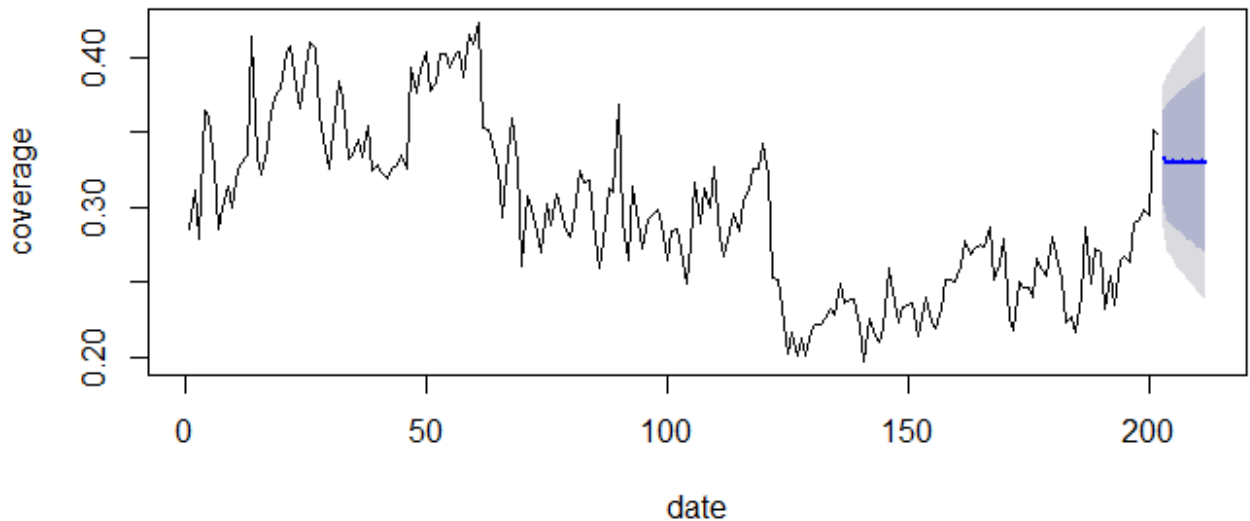


Рисунок 3.24 – Прогноз на тестовій вибірці за моделлю ARIMA для заповнюваності

У таблиці 3.9 наведено результати обчислення критеріїв, а в таблиці 3.10 наведено результати обчислення похибок.

Таблиця 3.9 – Значення критеріїв для доходу та заповнюваності

Критерій	Дохід	Заповнюваність
<i>AIC</i>	3202.2	-919.87
<i>BIC</i>	3222.01	-903.35
$R^2$	0.7522077	0.818524
<i>DW</i>	1.981691	2.006298

Таблиця 3.10 – Значення похибок для доходу та заповнюваності

Похибка	Дохід	Заповнюваність
<i>RMSE</i>	673.1706	0.02387146
<i>MAE</i>	536.188	0.01785158
<i>MAPE</i>	13.8156	6.081153
<i>MASE</i>	0.8847556	0.9483528

### 3.5.5 Побудова моделі ARIMAX

Насамкінець, потрібно побудувати модель ARIMAX. Як завжди буде виконана побудова двох моделей – для доходу та заповнюваності. Після цього наведемо значення показників адекватності, інформативності моделей, а також залишки моделей на навчальній вибірці та значення похибок і прогноз та тестовій вибірці.

Для доходу отримано модель ARIMAX(2, 1, 3, 3), а для заповнюваності – ARIMAX(1, 1, 3, 3).

Результати моделювання наведемо на рисунках 3.25-3.28.

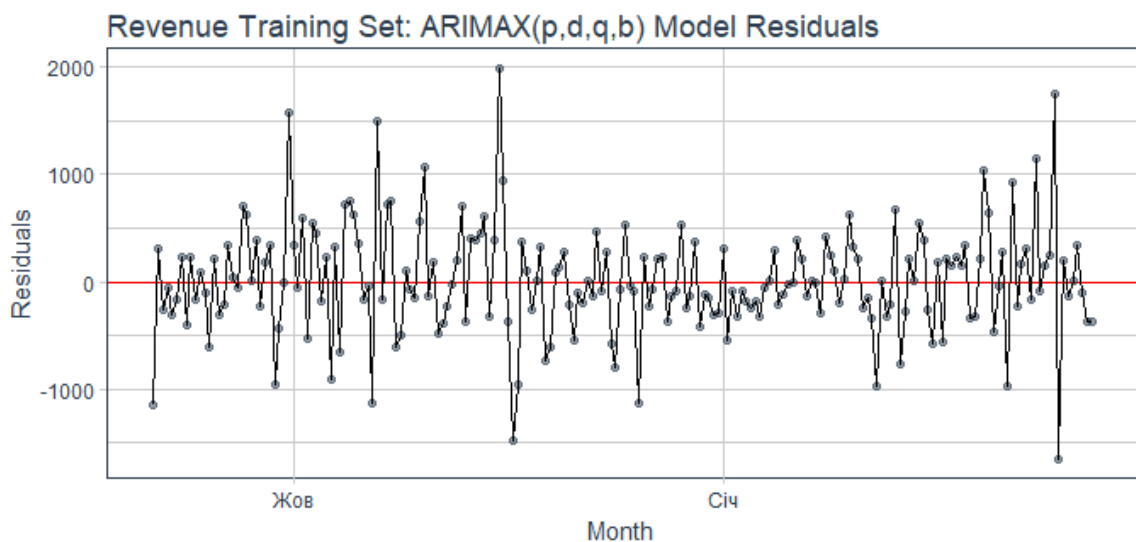


Рисунок 3.25 – Залишки для моделі ARIMAX для доходу

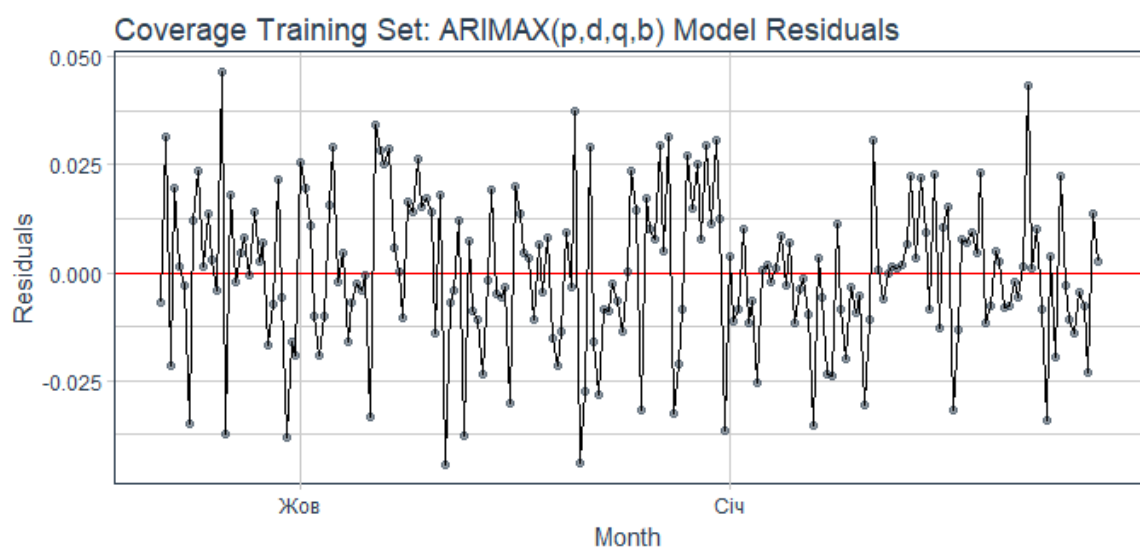


Рисунок 3.26 – Залишки для моделі ARIMAX для заповнюваності

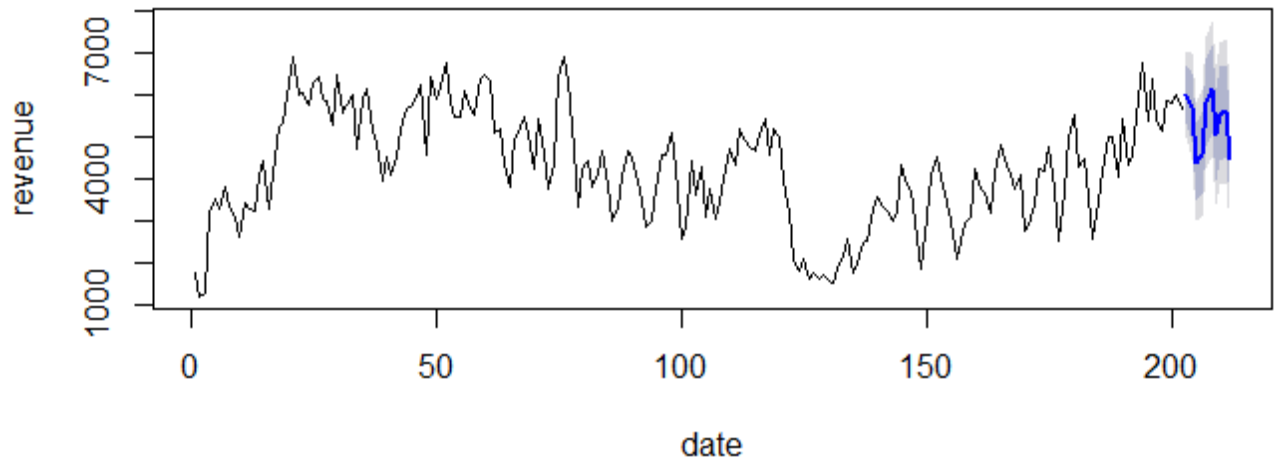


Рисунок 3.27 – Прогноз на тестовій вибірці за моделлю ARIMAX для доходу

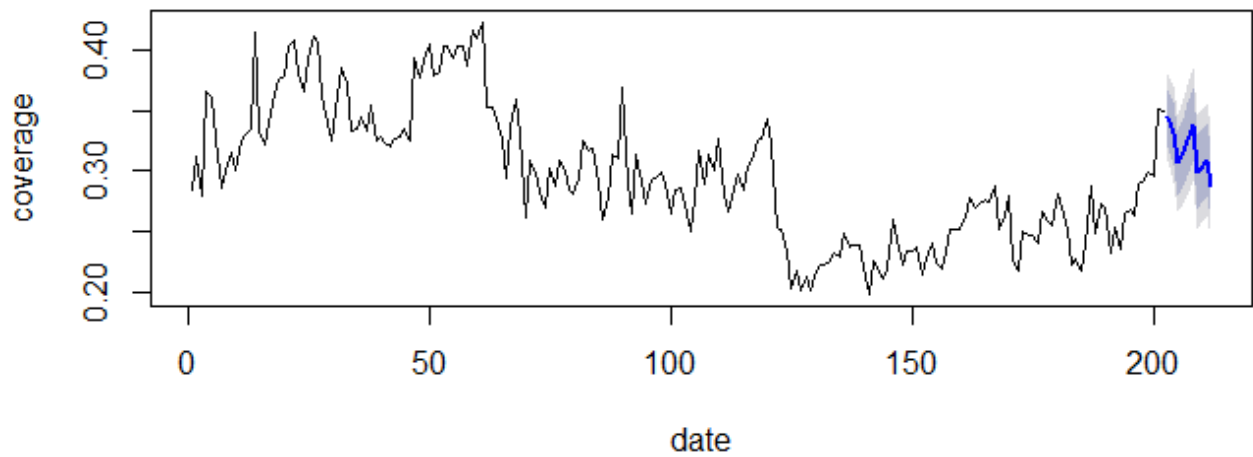


Рисунок 3.28 – Прогноз на тестовій вибірці за моделлю ARIMAX для заповнюваності

У таблиці 3.11 наведено результати обчислення критеріїв, а в таблиці 3.12 наведено результати обчислення похибок.

Таблиця 3.11 – Значення критеріїв для доходу та заповнюваності

Критерій	Дохід	Заповнюваність
<i>AIC</i>	3118.26	-1025.5
<i>BIC</i>	3161.27	-972.56
$R^2$	0.8531828	0.9000842
<i>DW</i>	2.007797	1.985912

Таблиця 3.12 – Значення похибок для доходу та заповнюваності

Похибка	Дохід	Заповнюваність
<i>RMSE</i>	509.021	0.01760031
<i>MAE</i>	373.7665	0.01380587
<i>MAPE</i>	9.315525	4.715555
<i>MASE</i>	0.6167463	0.7334275

### 3.6 Вибір кращої моделі прогнозу для доходу та заповнюваності

Отже, ми побудували всі бажані моделі. Тепер прийшов час виконати порівняльний аналіз побудованих моделей та вибрати найкращу з них для подальшого використання.

Для виконання порівняльного аналізу, створимо таблиці 3.13-3.16.

Таблиця 3.13 – Порівняння побудованих моделей для доходу за критеріями

Model	DW	$R^2$	AIC	BIC
AR	1.981992	0.714015	3233.09	3243.01
ARMA	1.981992	0.714015	3233.09	3243.01
ARIMA	1.981691	0.7522077	3202.2	3222.01
ARIMAX	2.007797	0.8531828	3118.26	3161.27

Таблиця 3.14 – Порівняння побудованих моделей для доходу за похибками

Model	RMSE	MAE	MAPE	MASE
AR	710.3305	587.1062	15.69017	0.9687749
ARMA	710.3305	587.1062	15.69017	0.9687749
ARIMA	673.1706	536.188	13.8156	0.8847556
ARIMAX	509.021	373.7665	9.315525	0.6167463

Таблиця 3.15 – Порівняння побудованих моделей для заповнюваності за критеріями

Model	DW	$R^2$	AIC	BIC
AR	1.993624	0.8180255	-924.14	-904.29
ARMA	2.003657	0.8186622	-922.83	-899.68
ARIMA	2.006298	0.818524	-919.87	-903.35
ARIMAX	2.001912	0.9000842	-1025.5	-972.56

Таблиця 3.16 – Порівняння побудованих моделей для заповнюваності за похибками

Model	RMSE	MAE	MAPE	MASE
AR	0.02374302	0.01780827	6.044972	0.9460523
ARMA	0.02370153	0.0177796	6.045356	0.9445291
ARIMA	0.02387146	0.01785158	6.081153	0.9483528
ARIMAX	0.01760031	0.01380587	4.715555	0.7334275

Таким чином, проаналізувавши та порівнявши таблиці 3.13-3.16, можна зробити висновок, що для розв'язання поставленої задачі слід використовувати модель ARIMAX, адже саме вона показала найкращі результати з точки зору критеріїв та похибок.

Насамкінець, потрібно побудувати прогнози на 10 кроків вперед. Проте для більшої точності ми побудуємо модель ARIMAX взявши до розгляду весь набір даних, тобто додамо 10 тестових значень до тренувальної вибірки.

Результати моделювання наведемо на рисунках 3.29-3.32.

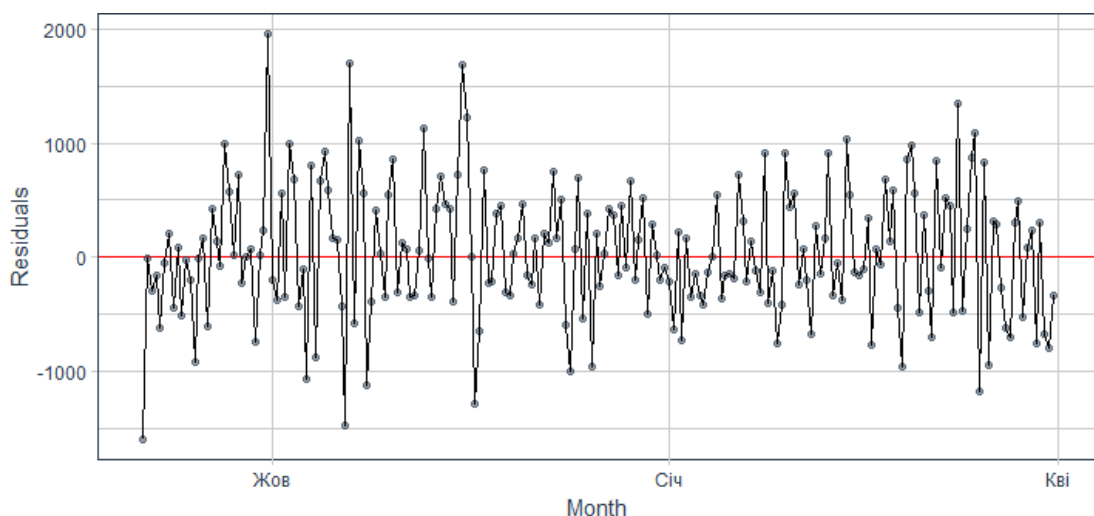


Рисунок 3.29 – Залишки на повній вибірці для моделі ARIMAX для доходу

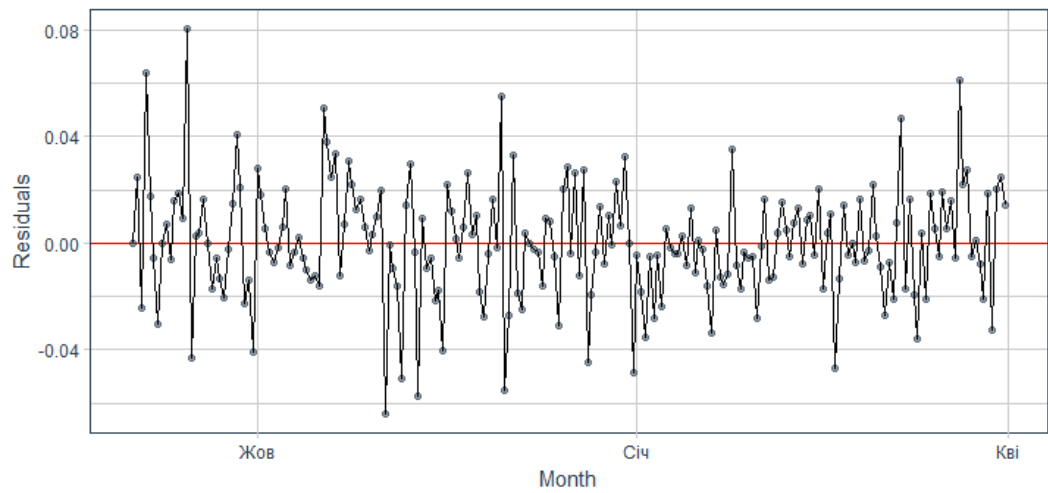


Рисунок 3.30 – Залишки на повній вибірці для моделі ARIMAX для заповнюваності

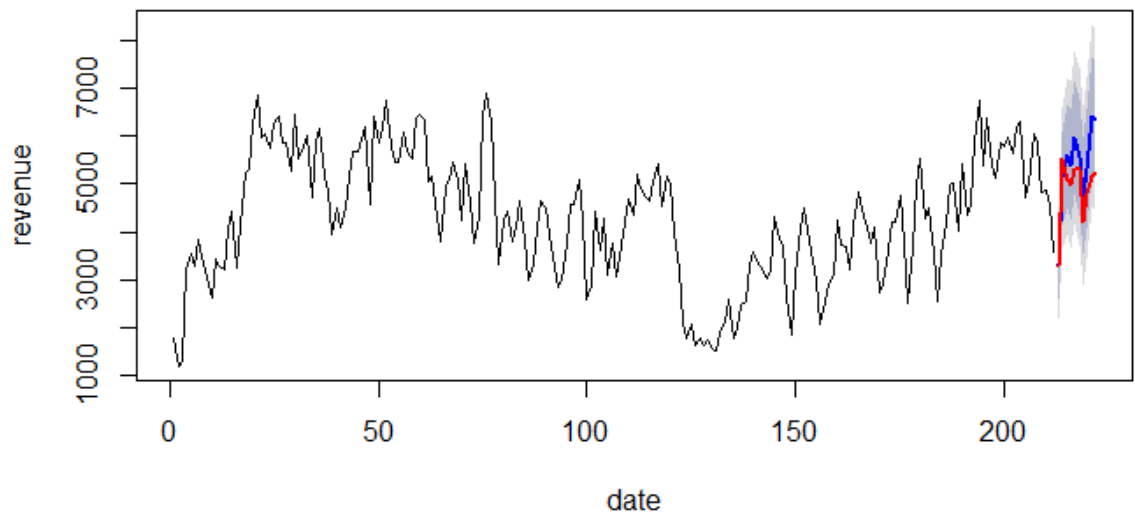


Рисунок 3.31 – Прогноз на 10 кроків вперед за моделлю ARIMAX для доходу

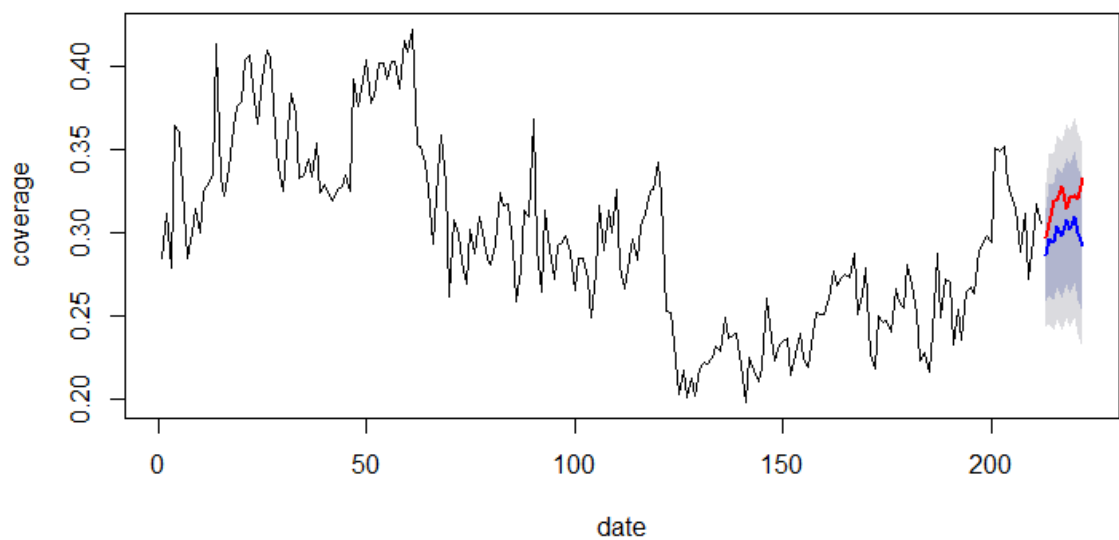


Рисунок 3.32 – Прогноз на 10 кроків вперед за моделлю ARIMAX для заповнюваності

У таблиці 3.17 наведено результати обчислення критеріїв, а в таблиці 3.18 наведено результати обчислення похибок.

Таблиця 3.17 – Значення критеріїв для доходу та заповнюваності

<b>Критерій</b>	<b>Дохід</b>	<b>Заповнюваність</b>
<i>AIC</i>	3313.06	-1005
<i>BIC</i>	3329.84	-971.44
$R^2$	0.8073593	0.8447321
<i>DW</i>	2.114268	1.992115

Таблиця 3.18 – Значення похибок для доходу та заповнюваності

<b>Похибка</b>	<b>Дохід</b>	<b>Заповнюваність</b>
<i>RMSE</i>	583.4133	0.02152746
<i>MAE</i>	458.3663	0.01609535
<i>MAPE</i>	11.48008	5.474169
<i>MASE</i>	0.756473	0.8554458

### Висновки до розділу 3

Отже, в даному розділі було проведено моделювання з метою побудови прогнозу ключових характеристик трафіка цифрової реклами.

Серед побудованих моделей були наступні: множинна регресія, модель авторегресії, модель авторегресії та ковзного середнього, модель авторегресії та інтегрованого ковзного середнього, модель авторегресії та інтегрованого ковзного середнього з екзогенними змінними.

Було проведено процес моделювання для двох ключових параметрів трафіка цифрової реклами. А саме – для доходу та показника заповнюваності.



При цьому також було проведено порівняльний аналіз побудованих моделей. В результаті даного порівняльного аналізу було виявлено, що найкращою моделлю, як для доходу, так і для показника заповнюваності, є модель авторегресії та інтегрованого ковзного середнього з екзогенними змінними.

Таким чином, для подальшого практичного використання слід застосовувати модель авторегресії та інтегрованого ковзного середнього з екзогенними змінними, оскільки для розв'язання поставленої задачі вона виявилась кращою за інші побудовані моделі.

## РОЗДІЛ 4 РОЗРОБЛЕННЯ СТАРТАП-ПРОЕКТУ

### 4.1 Опис ідеї проекту

У даному розділі наведемо таблиці 4.1-4.22, які описують розробку стартап-проекту [30].

Таблиця 4.1 – Опис ідеї стартап-проекту

<i>Зміст ідеї</i>	<i>Напрямки застосування</i>	<i>Вигоди для користувача</i>
Ідея стартапу – агрегатор рекламної статистики із функціями прогнозування найважливіших параметрів трафіка цифрової реклами. А саме – доходу та заповнюваності.	1. Швидкий збір статистики від великої кількості рекламодавців	Суттєве зменшення часу, який потрібен для збору та попередньої обробки статистики, яка надходить від великої кількості рекламодавців
	2. Можливість агрегування статистики	Зручне агрегування статистики по заданим параметрам, що дозволяє робити попередні висновки щодо зібраної статистики
	3. Можливість швидкого реагування на зміну тенденцій на ринку завдяки функції прогнозування важливих параметрів	Використання прогнозу доходу та заповнюваності для подальшого планування витрат та стратегічного планування подальшої роботи по певним напрямкам

У наступній таблиці виберемо бальну шкалу для оцінювання від 1 до 20.

Таблиця 4.2 – Визначення сильних, слабких та нейтральних характеристик ідеї проекту

№ n/n	Техніко- економічні характеристи ки ідеї	(потенційні) товари/концепції конкурентів			W (слабка сторона)	N (нейтральн а сторона)	S (сильна сторона)
		Мій проект	Конкуре нт 1	Конкуре нт 2			
1.	Функціональ ність	15	10	8	—	—	Спрощенн я збору статистики
2.	Складність виконання проекту	15	16	13	Складність виконання наукоємних завдань	—	—
3.	Наявність фінансових ресурсів	9	12	5	Недостатні фінансові ресурси	—	—
4.	Ціна продукту	12	9	10	—	Прийнятна ціна продукту	—
5.	Якість продукту	13	10	8	—	—	Висока якість продукту
6.	Простота збору статистики	15	10	8	—	—	Вихід на нові ринки збуту

## 4.2 Технологічний аудит ідеї проекту

Таблиця 4.3 – Технологічна здійсненність ідеї проекту

№ n/n	Ідея проекту	Технології її реалізації	Наявність технологій	Доступність технологій
1	Ідея стартапу – агрегатор рекламної статистики із функціями прогнозування найважливіших параметрів трафіка цифрової реклами. А саме – доходу та заповнюваності.	Розробка агрегатора статистики із використанням мови програмування PHP та бази даних MySQL	Дана технологія наявна на ринку	Дана технологія частково доступна автору проекту
2		Розробка агрегатора статистики із використанням мови програмування Python та бази даних MySQL	Дана технологія наявна на ринку	Дана технологія доступна автору проекту
3		Розробка агрегатора статистики із використанням мови програмування Python та бази даних PostgreSQL	Дана технологія наявна на ринку	Дана технологія частково доступна автору проекту
Обрана технологія реалізації ідеї проекту: Отже, із наведеної вище таблиці, можна зробити висновок, що потрібно використовувати для розробки продукту мову програмування Python та базу даних MySQL.				

## 4.3 Аналіз ринкових можливостей запуску стартап-проекту

Таблиця 4.4 – Попередня характеристика потенційного ринку стартап проекту

<i>№ n/n</i>	<i>Показники стану ринку (найменування)</i>	<i>Характеристика</i>
1	Кількість головних гравців, од	3
2	Загальний обсяг продаж, грн / ум. од	---
3	Динаміка ринку (якісна оцінка)	Зростає
4	Наявність обмежень для входу (вказати характер обмежень)	Обмежень немає
5	Специфічні вимоги до стандартизації та сертифікації	Відсутні
6	Середня норма рентабельності в галузі (або по ринку), %	25%

Таблиця 4.5 – Характеристика потенційних клієнтів стартап-проекту

<i>№ n/n</i>	<i>Потреба, що формує ринок</i>	<i>Цільова аудиторія (цільові сегменти ринку)</i>	<i>Відмінності у поведінці різних потенційних цільових груп клієнтів</i>	<i>Вимоги споживачів до товару</i>
1	Ринок потребує агрегатор статистики, який буду спрощувати її збір.	Компанії, що займаються цифровою рекламою	Кожна компанія має свій набір специфічних вимог, які наперед передбачити неможливо. Тому продукт має бути універсальним	Швидкодія роботи програми
2				Функціональність
3				Наявність великої кількості рекламодавців

Таблиця 4.6 – Фактори загроз

<i>№ n/n</i>	<i>Фактор</i>	<i>Зміст загрози</i>	<i>Можлива реакція компанії</i>
1	Нерентабельність	Затратність розробки	Зниження собівартості
2	Підвищення конкуренції	Наявність аналогів	Збільшення функціоналу

Таблиця 4.7 – Фактори можливостей

<i>№ n/n</i>	<i>Фактор</i>	<i>Зміст можливості</i>	<i>Можлива реакція компанії</i>
1	Новизна	Новий продукт на ринку	Постійний аналіз ринку
2	Прибутковість	Підвищення прибутковості	Подальший розвиток

Таблиця 4.8 – Ступеневий аналіз конкуренції на ринку

<i>Особливості конкурентного середовища</i>	<i>В чому проявляється дана характеристика</i>	<i>Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)</i>
1. Вказати тип конкуренції - монополія/олігополія/ монополістична/чиста	Олігополія	Покращення функціоналу
2. За рівнем конкурентної боротьби - локальний/національний/...	Локальний	Глобалізація
3. За галузевою ознакою - міжгалузева/ внутрішньогалузева	Внутрішньогалузева	Покращення позицій в галузі
4. Конкуренція за видами товарів: - товарно-родова - товарно-видова - між бажаннями	Між бажаннями	Генерація нових ідей
5. За характером конкурентних переваг - цінова / нецінова	Нецінова	Покращення якості
6. За інтенсивністю - марочна/не марочна	Не марочна	Підвищення відомості бренду

Таблиця 4.9 – Аналіз конкуренції в галузі за М. Портером

Складові аналізу	<i>Прямі конкуренти в галузі</i>	<i>Потенційні конкуренти</i>	<i>Постачальники</i>	<i>Клієнти</i>	<i>Товари-замінники</i>
	<i>Admixer</i>	<i>Початковий капітал</i>	<i>Відсутні</i>	<i>Внутрішній продукт</i>	<i>Агрегатори-аналоги</i>
Висновки:	Інтенсивність конкуренції невисока	Конкурентів небагато. Строки виходу на ринок – 12 місяців	---	Клієнти не диктують умови	Вихід на глобальний ринок потрібно досліджувати додатково

Таблиця 4.10 – Обґрунтування факторів конкурентоспроможності

<i>№ n/n</i>	<i>Фактор конкурентоспроможності</i>	<i>Обґрунтування (наведення чинників, що роблять фактор для порівняння конкурентних проектів значущим)</i>
1	Функціональність	Достатній функціонал для збору статистики

Таблиця 4.11 – Порівняльний аналіз сильних та слабких сторін «агрегатора статистики»

<i>№ n/n</i>	<i>Фактор конкурентоспроможності</i>	<i>Бали 1-20</i>	<i>Рейтинг товарів-конкурентів у порівнянні з Mediawayss</i>						
			-3	-2	-1	0	+1	+2	+3
1	Функціональність	15			+				
2	Складність виконання проекту	15			+				
3	Наявність фінансових ресурсів	9					+		
4	Ціна продукту	12				+			
5	Якість продукту	13			+				
6	Простота збору статистики	15			+				

Таблиця 4.11 – SWOT-аналіз стартап-проекту

<p>Сильні сторони:</p> <ul style="list-style-type: none"> <li>- Спрощення збору статистики</li> <li>- Прийнятна ціна продукту</li> <li>- Висока якість продукту</li> </ul>	<p>Слабкі сторони:</p> <ul style="list-style-type: none"> <li>- Недостатні фінансові ресурси</li> <li>- Складність виконання наукоємних завдань</li> </ul>
<p>Можливості:</p> <ul style="list-style-type: none"> <li>- Підвищення прибутковості</li> <li>- Вихід на нові ринки збуту</li> </ul>	<p>Загрози:</p> <ul style="list-style-type: none"> <li>- Підвищення конкуренції</li> <li>- Зміна потреб клієнтів</li> <li>- Брак коштів у клієнтів</li> </ul>

Таблиця 4.13 – Альтернативи ринкового впровадження стартап-проекту

<i>№ n/n</i>	<i>Альтернатива (орієнтовний комплекс заходів) ринкової поведінки</i>	<i>Ймовірність отримання ресурсів</i>	<i>Строки реалізації</i>
1	Внутрішній продукт	100%	12 місяців

#### 4.4 Розроблення ринкової стратегії проекту

Таблиця 4.124 – Вибір цільових груп потенційних споживачів

<i>№ n/n</i>	<i>Опис профілю цільової групи потенційних клієнтів</i>	<i>Готовність споживачів сприйняти продукт</i>	<i>Орієнтовний попит в межах цільової групи (сегменту)</i>	<i>Інтенсивність конкуренції в сегменті</i>	<i>Простота входу у сегмент</i>
1	Компанії, що займаються цифровою рекламою	Споживачі готові сприйняти продукт	Попит знаходиться на середньому рівні	Конкуренція незначна	Увійти в сегмент досить просто
Які цільові групи обрано: компанії, що займаються цифровою рекламою					

Таблиця 4.15 – Визначення базової стратегії розвитку

<i>№ n/n</i>	<i>Обрана альтернатива розвитку проекту</i>	<i>Стратегія охоплення ринку</i>	<i>Ключові конкурентоспроможні позиції відповідно до обраної альтернативи</i>	<i>Базова стратегія розвитку</i>
1	Розробка агрегатора статистики як внутрішнього продукту з подальшою можливістю виводу на ринок	Пропонування продукту компаніям-партнерам	Невеликі початкові затрати Можливість реального тестування до виводу на ринок	Стратегія диференціації

Таблиця 4.16 – Визначення базової стратегії конкурентної поведінки

<i>№ n/n</i>	<i>Чи є проект «першопрохідцем» на ринку?</i>	<i>Чи буде компанія шукати нових споживачів, або забирати існуючих у конкурентів?</i>	<i>Чи буде компанія копіювати основні характеристики товару конкурента, і які?</i>	<i>Стратегія конкурентної поведінки</i>
1	Ні, проект є вдосконаленням вже існуючих підходів до збору статистики	Компанія буде шукати нових споживачів	Компанія не буде копіювати основні характеристики товару конкурента	Стратегія виклику лідера



Таблиця 4.17 – Визначення стратегії позиціонування

<i>№ n/n</i>	<i>Вимоги до товару цільової аудиторії</i>	<i>Базова стратегія розвитку</i>	<i>Ключові конкурентоспро- можні позиції власного стартап-проекту</i>	<i>Вибір асоціацій, які мають сформувані комплексну позицію власного проекту (три ключових)</i>
1	Швидкодія роботи програми	Наступальна стратегія	Висока швидкодія	Висока швидкість Відсутність зависань Наявність можливості повторного виконання запитів
2	Функціональність		Просунута функціональність	Надійність Багато функцій Якість
3	Наявність великої кількості рекламодавців		Значна кількість рекламодавців	Багато рекламодавців Гнучкість Масштабованість

#### 4.5 Розроблення маркетингової програми стартап-проекту

Таблиця 4.18 – Визначення ключових переваг концепції потенційного товару

<i>№ n/n</i>	<i>Потреба</i>	<i>Вигода, яку пропонує товар</i>	<i>Ключові переваги перед конкурентами (існуючі або такі, що потрібно створити)</i>
1	Швидкодія	Висока швидкодія	Збільшення швидкості збору статистики
2	Функціональність	Висока функціональність	Надання широкої функціональності
3	Велика кількість рекламодавців	Велика база рекламодавців	Збільшення кількості рекламодавців

Таблиця 4.19 – Опис трьох рівнів моделі товару

<i>Рівні товару</i>	<i>Сутність та складові</i>
I. Товар за задумом	- Швидкий збір необхідної статистики - Агрегування статистики для її подальшого аналізу
II. Товар у реальному виконанні	Властивості / характеристики
	1. Функціональність
	2. Швидкодія
	3. Велика база рекламодавців
	Назва організації-розробника: Mediawayss
III. Товар із підкріпленням	До продажу: різноманітні знижки на придбання продукту
	Після продажу: гарантія якості, подальша підтримка продукту
За рахунок чого потенційний товар буде захищено від копіювання: За рахунок прив'язки програми до особистого акаунту на сайті компанії-виробника.	

Таблиця 4.20 – Визначення меж встановлення ціни

<i>№ п/п</i>	<i>Рівень цін на товари-замінники</i>	<i>Рівень цін на товари-аналоги</i>	<i>Рівень доходів цільової групи споживачів</i>	<i>Верхня та нижня межі встановлення ціни на товар/послугу</i>
1	\$ 2000	\$ 10000	\$ 200000	\$ 8000 - \$ 12000

Таблиця 4.21 – Формування системи збуту

<i>№ п/п</i>	<i>Специфіка закупівельної поведінки цільових клієнтів</i>	<i>Функції збуту, які має виконувати постачальник товару</i>	<i>Глибина каналу збуту</i>	<i>Оптимальна система збуту</i>
1	Одноразова купівля програмного продукту з його подальшою підтримкою	Встановлення контактів із споживачами і підтримання їх; Формування попиту і стимулювання збуту; Організація руху товару.	0 - 1	Безпосередньо від виробника до споживача

Таблиця 4.22 – Концепція маркетингових комунікацій

<i>№ n/n</i>	<i>Специфіка поведінки цільових клієнтів</i>	<i>Канали комунікацій, якими користуються цільові клієнти</i>	<i>Ключові позиції, обрані для позиціонування</i>	<i>Завдання рекламного повідомлення</i>	<i>Концепція рекламного звернення</i>
1	Найбільший рівень довіри проявляють до партнерів	Взаємодія з компаніями- партнерами	Функціонал  Ціна  Швидкодія	Розповісти споживачам про наявність даного продукту та його потенційну користь	Контент- маркетинг

## Висновки до розділу 4

Отже, в результаті виконання даного розділу магістерської дисертації, можна зробити висновок, що існує реальна можливість ринкової комерціалізації проекту. При цьому, швидше за все буде попит, динаміка ринку є позитивною. Також варто зазначити, що такий вид діяльності на ринку буде рентабельним.

З огляду на потенційні групи клієнтів та бар'єри, які стоять на шляху, можна сказати, що у даного проекту є досить непогані перспективи впровадження. При цьому рівень конкуренції на даний момент є не дуже високим, а конкурентоспроможність проекту є достатньою.

Для ринкової реалізації проекту, на даний момент, краще обрати варіант розробки продукту, при якому використовується мова програмування Python та база даних MySQL.

З огляду на все сказане вище, можна підсумувати, що подальша імплементація проекту є доцільною. Тому його можна виводити на ринок, та отримувати при цьому прибуток.

## ВИСНОВКИ ЗА РОБОТОЮ ТА ПЕРСПЕКТИВИ ДЛЯ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Магістерська дисертація присвячена задачі прогнозування найважливіших характеристик трафіка цифрової реклами. А саме – описано процес побудови моделей прогнозу цих характеристик та вибору кращої моделі за точністю.

У даній роботі описано поняття цифрової реклами, її трансформацію з часом, а також – типи та види. При цьому, було наведено актуальність даної тематики та пояснено необхідність прогнозування найважливіших характеристик цифрового рекламного трафіка.

Здійснено загальний огляд методів моделювання та прогнозування статистичних даних у формі часових рядів. Після цього були вибрані найбільш підходящі методи, що відповідають поставленим задачам моделювання та прогнозування.

Виконано вибір інструментальної платформи для здійснення прогнозування, що відповідає поставленим вимогам. А саме, було дві мови програмування – R та Python. Вони разом дозволяють здійснювати якісний збір даних, їхню очистку та підготовку, а також – власне кажучи процедури моделювання та прогнозування.

Розроблено схему розв'язку поставленої задачі прогнозування найважливіших характеристик трафіка цифрової реклами. Зокрема наведено опис побудови моделей часових рядів для прогнозування конкретних параметрів цифрового рекламного трафіка. Розроблену схему застосовано до набору реальних статистичних даних, що описують трафік цифрової реклами деякої рекламної компанії-посередника, та проведений порівняльний аналіз побудованих 4 моделей:

- *AR*;
- *ARMA*;
- *ARIMA*;
- *ARIMAX*.

В результаті порівняння розроблених моделей за допомогою інформаційних критеріїв та похибок, виявилось, що для розв'язання поставленої задачі найкращою є модель авторегресії та інтегрованого ковзного середнього з екзогенними змінними. При використанні даної моделі спостерігається значення похибки *MAPE* в інтервалі  $[0; 10]$ , що свідчить про високу точність моделі.

Таким чином, ми з'ясували, що для прогнозування характеристик трафіка цифрової реклами, найбільш підходящою є модель *ARIMAX*. Тому саме цю модель варто використовувати на практиці для побудови прогнозів ключових характеристик трафіка цифрової реклами.

Новизна отриманих результатів полягає у тому, що на основі отриманого набору статистичних даних, було побудовано нові моделі часових рядів: *AR*, *ARMA*, *ARIMA*, *ARIMAX*. При цьому, порівняльний аналіз побудованих моделей показав, що найкращою є модель *ARIMAX*, тому саме цю модель потрібно використовувати для подальшого практичного використання.

На основі магістерської дисертації були написані тези, а також написана наукова стаття. Тези будуть опубліковані в збірці тез доповідей конференції САІТ-2018. А наукова стаття буде опублікована в електронній збірці доповідей у видавництві CEUR.

Також варто зазначити, що за даною темою існує перспектива для подальших досліджень, оскільки галузь цифрової реклами є досить популярною у наш час, і при цьому досить стрімко розвивається в останні роки. Зокрема, можна побудувати нові прогностичні моделі, наприклад: застосувати експоненційне згладжування, а потім підібрати модель авторегресії та тренду (*ARX*); використати метод МГУА для однієї змінної (*AR* – МГУА). Також одним із перспективних шляхів подальшого вдосконалення розроблених підходів, є прогнозування ключових характеристик не для всього трафіка цифрової реклами, а для окремих найбільш значущих сайтів із цього трафіка. Ще одним можливим шляхом розвитку даного дослідження, могло б стати застосування розроблених підходів до прогнозування до нових наборів даних, з метою перевірки коректності та адекватності такого підходу.

## ПЕРЕЛІК ПОСИЛАНЬ

1. Ryan D. Digital Marketing For Dummies / D. Ryan, H. Russ. – New Jersey: John Wiley & Sons, 2017. – 360 p.
2. Ryan D. Understanding digital marketing: marketing strategies for engaging the digital generation / D. Ryan, C. Jones. – New York: Kogan Page, 2017. – 276 p.
3. Micael D. Marketing Communications: A Brand Narrative Approach / D. Micael, L. Fredrik, S. Terry. – Chichester: John Wiley & Sons, 2010. – 606 p.
4. Understanding consumer database marketing / [D. D. Schoenbachler, G. L. Gordon, D. Foley, L. Spellman]. // Journal of Consumer Marketing. – 1997. – №1. – pp. 5–19.
5. Hart C. Retailer adoption of the Internet – Implications for retail marketing / C. Hart, N. Doherty, F. Ellis-Chadwick. // European Journal of Marketing. – 2000. – №8. – pp. 954–974.
6. Kates M. Making digital and traditional marketing work together [Електронний ресурс] / Matthew Kates // Econsultancy. – 2013. – Режим доступу до ресурсу: <http://econsultancy.com/au/blog/62546-making-digital-and-traditional-marketing-work-together>
7. Digital Marketing For Dummies / [B.Carter, G. Brooks, F. Catalano, B. E. Smith]. – Chichester: John Wiley & Sons, 2007. – 304 p.
8. McCambley J. The first ever banner ad: why did it work so well? [Електронний ресурс] / Joe McCambley // The Guardian. – 2013. – Режим доступу до ресурсу: <https://www.theguardian.com/media-network/media-network-blog/2013/dec/12/first-ever-banner-ad-advertising>.
9. Worldwide Ad Growth Buoyed by Digital, Mobile Adoption [Електронний ресурс] // eMarketer. – 2013. – Режим доступу до ресурсу: <https://www.emarketer.com/Article/Worldwide-Ad-Growth-Buoyed-by-Digital-Mobile-Adoption/1010244>.
10. Digital Marketing Communication [Електронний ресурс] // International Chamber of Commerce. – 2010. – Режим доступу до ресурсу:

<https://iccwbo.org/global-issues-trends/responsible-business/marketing-advertising/digital-marketing-communication/>.

11. What Are Connected Shoppers Doing And Not Doing Online? [Электронный ресурс] // Nielsen Global. – 2016. – Режим доступа до ресурсу: <http://www.nielsen.com/us/en/insights/news/2016/what-are-connected-shoppers-doing-and-not-doing-online.html>.
12. Key Success Factors In Digital Marketing In Service Industry and the Development Strategies: A Case Study On Fleur DE Chine At Sun Moon Lake / [S.Maw-Liann, C. Wan-Ju, C. Wen-Yuan, W. Sheng-Liang]. // The international Journal of organizational innovation. – 2015. – №1. – pp. 171–185.
13. Connected Commerce Is Creating Buyers Without Borders [Электронный ресурс] // Nielsen Global. – 2016. – Режим доступа до ресурсу: <http://www.nielsen.com/us/en/insights/news/2016/connected-commerce-is-creating-buyers-without-borders.html>.
14. Mogos R. I. Digital Marketing for Identifying Customers' Preferences -- A Solution for SMEs in Obtaining Competitive Advantages / Mogos. // International Journal of Economic Practices & Theories. – 2015. – №3. – pp. 240–247.
15. From Social to Sale: The Effects of Firm-Generated Content in Social Media on Customer Behavior / [A. Kumar, R. Bezawada, R. Rishika and other]. // Journal of Marketing. – 2016. – №1. – pp. 7–25.
16. Van Niekerk A. Strategic management of media assets for optimizing market communication strategies, obtaining a sustainable competitive advantage and maximizing return on investment: An empirical study / Van Niekerk. // Journal of Digital Asset Management. – 2007. – №2. – pp. 89–98
17. The influence of social media interactions on consumer–brand relationships: A three-country study of brand perceptions and marketing behaviors / [S.Hudson, L. Huang, M. S. Roth, T. J. Madden]. // International Journal of Research in Marketing. – 2016. – №1. – pp. 27–47.

18. Digital advertising revenue of leading online companies 2012-2017 [Електронний ресурс] // Statista – The Statistics Portal. – 2017. – Режим доступу до ресурсу: <https://www.statista.com/statistics/205352/digital-advertising-revenue-of-leading-online-companies/>.
19. Box G. Time Series Analysis: Forecasting & Control / G. Box, G. M. Jenkins, G. Reinsel. – [4th edition]. – New Jersey: John Wiley & Sons, 2008. – 784 p.
20. Wei W. W. S. Time Series Analysis: Univariate and Multivariate Methods / William Wei. – New York: Pearson, 2005. – 624 p.
21. Brockwell P. J. Time Series: Theory and Methods / P. J. Brockwell, R. A. Davis. – New York: Springer, 2009. – 580 p.
22. Hyndman R. J. Forecasting: principles and practice / R. J. Hyndman, G. Athanasopoulos. – Melbourne: OTexts, 2013. – 292 p.
23. Makridakis S. G. Forecasting: Methods and Applications / S. G. Makridakis, S. C. Wheelwright, R. J. Hyndman. – New York: John Wiley & Sons, 2011. – 656 p.
24. Armstrong S. J. Principles of Forecasting: A Handbook for Researchers and Practitioners / Armstrong. – New York: Springer, 2002. – 850 p.
25. Бідюк П. І. Аналіз часових рядів (навчальний посібник) / П. І. Бідюк, В. Д. Романенко, О. Л. Тимошук. – Київ: Політехніка, 2010. – 317 с.
26. Hamilton J. D. Time Series Analysis / James Douglas Hamilton. – Princeton: Princeton University Press, 1994. – 820 p.
27. Bisgaard S. Time Series Analysis and Forecasting by Example / S. Bisgaard, M. Kulahci. – New Jersey: John Wiley & Sons, 2011. – 400 p.
28. Shumway R. H. Time Series Analysis and Its Applications: With R Examples / R. H. Shumway, D. S. Stoffer. – New York: Springer, 2011. – 596 p.
29. Cryer J. D. Time Series Analysis: With Applications in R / J. D. Cryer, K. Chan. – New York: Springer, 2010. – 491 p.
30. Розроблення стартап-проекту [Електронний ресурс]: Методичні рекомендації до виконання розділу магістерських дисертацій для студентів інженерних спеціальностей / За заг. ред. О.А. Гавриша. – Київ : НТУУ «КПІ», 2016. – 28 с.